



L OVELY
P ROFESSIONAL
U NIVERSITY

**SUMMARIZING CUSTOMER REVIEWS – A TEXT
SUMMARIZATION APPROACH**

A Dissertation

Submitted

By

Nidhi

To

Department of Computer Science and Engineering

In partial fulfillment of the Requirement for the

Award of the Degree of

Master of Technology in Computer Science and Engineering

Under the guidance of

Mr. Ravinder Singh

(Assistant Professor, Lovely Professional University)

(April 2015)

PAC APPROVAL



School of: Technology and Science

DISSERTATION TOPIC APPROVAL PERFORMA

Name of the Student: Nishi Registration No: 11306624
Batch: 2013-2015 Roll No: A25
Session: 2014 Parent Section: K2308
Details of Supervisor: Designation: Asst. Professor
Name: Ravinder Singh Qualification: Mtech
U.ID: 17750 Research Experience: 1yr

SPECIALIZATION AREA: DATABASE (pick from list of provided specialization areas by D.A.S.)

PROPOSED TOPICS

- Automatic Text Summarization (Opinion mining and checking & mining reviews and giving better algorithm in it content)
- Opinion Mining
- Improv. Efficiency of Mining algo.

Ravinder Singh
Signature of Supervisor
17750

PAC Remarks:

First topic is approved, publication expected.
1st ✓ 11/17/14

APPROVAL OF PAC CHAIRPERSON:

Signature: [Signature]
Date: 30/9/14

* Supervisor should finally encircle one topic out of three proposed topics and put up for a approval before Project Approval Committee (PAC)

* Original copy of this format after PAC approval will be retained by the student and must be attached in the Project/Dissertation final report.

* One copy to be submitted to Supervisor.

ABSTRACT

Reviews are unbiased information obtained from the sources outside an organization, which makes them more reliable in the eyes of customers. Online shoppers are very much concerned about product reviews before making any decision regarding buying the product. Product reviews play an important role in determining what kind of product is. Such reviews provide useful information about customer concern and their experience with the product. Consequently, these reviews will be helpful for a business making products for the purpose of product recommendation, better customer understanding and attracting more loyal customers. As ecommerce has become so popular, numbers of reviews are increasing day by day. It is difficult for a customer to read all the reviews manually. In this paper, an approach is developed which is used to obtain the summary from thousands or hundreds of online reviews. This approach uses extraction summarization for summarizing the reviews thereby selecting the original sentences and putting it together into a new shorter text explaining the overall opinion about the product. Although previous studies of deriving useful information from customer reviews focus on categorical or numerical data and textual data has been ignored. But textual data are of equal importance so it should not be ignored. So, this approach includes every aspect of the review in the summary so that a customer would be able to make a right decision regarding product.

ACKNOWLEDGEMENT

Gratitude cannot be seen or expressed. It can only be felt in heart and is beyond description. Often words are inadequate to serve as a model of expression of one's feeling, specially the sense of indebtedness and gratitude to all those who help us in our duty.

This report would not have been possible without the support of many people. Many thanks to my guide Mr. Ravinder Singh, who guided me thoroughly, read my numerous revisions and helped me by clearing out my doubts. Thanks to the Lovely Professional University, Department of CSE/IT for allowing me to pursue my dissertation in the topic of my choice.

And finally, thanks to my classmates, parents, and numerous friends who endured this long process with me, always offering support and love.

(Nidhi)

DECLARATION

I hereby declare that the dissertation-II proposal entitled **Summarizing Customer Reviews- A Text Summarization Approach** submitted for the M.Tech Degree is entirely my original work and all ideas and references have been duly acknowledged. It does not contain any work for the award of any other degree or diploma.

Date: _____

Investigator: Nidhi

Registration No. 11306674

CERTIFICATE

This is to certify that **Nidhi Sharma** has completed M.Tech dissertation proposal titled **Summarizing Customer Reviews- A Text Summarization Approach** under my guidance and supervision. To the best of my knowledge, the present work is the result of her original investigation and study. No part of the dissertation proposal has ever been submitted for any other degree or diploma. The dissertation proposal is fit for the submission and the partial fulfillment of the conditions for the award of M.Tech Computer Science & Engineering.

Date:

Signature of Advisor

Name: Ravinder Singh

UID: 17750

Table of Contents

CHAPTER No. CHAPTER NAME	PAGE No.
CHAPTER 1	1
1.1 Need of Customer Reviews	1
1.2 Processing of online customer reviews	2
1.3 Different Level of Analysis	3
1.3.1. Document level:	3
1.3.2. Sentence level:	4
1.3.3. Entity and Aspect level	4
1.4 Text Summarization Approach	4
1.4.1. Indicative summaries	4
1.4.2. Informative summaries	4
1.4.3. Critical summaries	4
1.5 Sentiment Analysis	5
1.5.1 Techniques used in sentiment analysis	7
1.5.2. Subtasks in Sentiment Analysis	8
1.5.3 Applications of sentiment analysis	8
1.5.4 Challenges in the sentiment analysis	9
CHAPTER 2 LITERATURE SURVEY	10
CHAPTER 3 PRESENT WORK	22
3.1 Problem Formulation	22
3.2 Objectives of the study	22
3.3 Research Methodology	24
3.3.1 WORKFLOW DIAGRAM	29

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

CHAPTER 4	RESULTS AND DISCUSSIONS	30
CHAPTER 5	CONCLUSIONS AND FUTURE SCOPE	457
CHAPTER 6	REFERENCES	49
CHAPTER 7	APPENDIX	51

LIST OF TABLES

Table No.:	Table Name	PAGE No.
Table 1:	Analysis of various numbers of messages	33
Table 2:	A table of statistical error calculated manually on the given database	33
Table 3:	A table of Precision and Recall for various densities of messages	34
Table 4:	The table of comparison between existing and proposed model	34

LIST OF FIGURES

FIGURE No.:	FIGURE NAME	PAGE No.
Figure1:	Sentiment Analysis	6
Figure2:	Flow diagram	28
Figure3:	Sentiment Analysis on 10 Messages	29
Figure4:	Sentiment Analysis on 100 Messages	30
Figure5:	Sentiment Analysis on 500 Messages	30
Figure6:	Sentiment Analysis on 1000 Messages	31
Figure7:	Sentiment Analysis on 2000 Messages	31
Figure8:	Sentiment Analysis on 5000 Messages	32
Figure9:	Sentiment Analysis on 10000 Messages	32
Figure10:	Sentiment Analysis Report of Nokia 920	36
Figure11:	Elapsed time per message of Nokia 920	36
Figure12:	Sentiment Analysis Report of Nokia1020	37
Figure13:	Elapsed time per message of Nokia 1020	37
Figure14:	Sentiment Analysis Report of Samsung S6 Edge	38
Figure15:	Elapsed time per message of Samsung S6 Edge	38
Figure16:	Sentiment Analysis Report of Samsung A5	39
Figure17:	Elapsed time per message of Samsung A5	40
Figure18:	Sentiment Analysis Report of Sony Xperia Z1	41
Figure19:	Elapsed time per message of Sony Xperia Z1	41
Figure20:	Sentiment Analysis Report of Sony Xperia M4	42
Figure21:	Elapsed time per message of Sony Xperia M4	42
Figure22:	Comparison between Nokia Lumia 920 and Samsung S6 Edge	43
Figure 23:	Elapsed time per message	44
Figure24:	Comparison between Samsung A5 and Sony Xperia Z1	44
Figure25:	Elapsed time per message	45
Figure26:	Comparison between Nokia Lumia 920 and Samsung S6 Edge on basis of Display feature	46

CHAPTER 1

INTRODUCTION

Product reviews play a vital role in a selection of a particular product. Customer reviews about a product are considered as sales drivers and are something that most of the customers will want to know before making a decision to buy a product. It is a fact that online customer reviews are trusted nearly twelve times more than the description provided by the manufacturers. Inventors of ecommerce like Amazon and eBay have been using product reviews since 1997, they lead people to write their opinion and share their experience about the products they have used so that it would help intend customers to buy that product or not.

1.1 Need of Customer Reviews

Gathering reviews from customers act as an asset for an organisation involved in selling products as this will help manufacturers to aware about strength and weakness of their products and help them to improve it. When going to buy product online, customers usually look at ratings of the product, read out reviews given by other customers and then compare the product with other products of same category. Quite simply, customer reviews increase conversions. Customer reviews help in improving the online businesses. Organizations look out the reviews given by customers to know what improvements they can make in their product. Different organizations use several different ways of getting product reviews from customers. For an instance, Amazon has a brilliant model for gaining product reviews through email. Social media like Facebook, Twitter and many other are considered as reliable sources of getting reviews. Using customer service or suggestion cards, customers are suggested to leave their thoughts and opinions about products. But it is difficult for a customer to go through hundreds or thousands of reviews in order to decide to buy a product or not. In response, this paper has proposed a technique for the summarization of customer reviews.

There are various reasons that show the importance of customer reviews for an organization selling products online. (<http://www.marketingdonut.co.uk>)

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

- i) Whenever a company introduces a new product then customer feedback is very important for deterring customer needs and tastes.
- ii) Companies can better understand that how their products are better than competitive products by analysing the customer ratings of product and their reasons for selection.
- iii) Companies can determine whether their customers are getting satisfactory level of service by their employees.
- iv) Customer reviews help in deterring why consumers are no longer interested in buying products from them, if any. This will help in building up strategies that would help lose customers back into business.
- v) Customer reviews are also important in determining technological trends in the market.

1.2 Processing of online customer reviews

The work done earlier in this field focused on classifying whole documents including overall positive or negative polarity as well as rating the scores of reviews.

Most of the existing methods work on processing the customer reviews online concentrates on opinion mining with the aim of knowing reviewer's attitude in terms of either positive or negative opinion towards various features of product. For an example shown below as Eg1, opinions regarding picture quality of digital camera are specified.

Digital Camera

Feature discussed over here is Picture Quality

Positive review sentences: 300

Negative review sentences: 30

Eg1: An example of opinion mining.

So in above example, quality of picture of digital camera is considered as good picture quality because more number of opinions is positive.

However there are some customer comments or reviews which cannot be described as positive or negative but certainly they are valuable. For an instance: Reviews regarding Nokia phone is stated as

1. The sound quality of Nokia phone is superb.
2. The most important thing for the review is sound quality.

Here, both the sentences are about the product feature that is sound quality. But 2nd sentence doesn't interpret any attitude orientation that is nether positive or negative. But it

expresses the customer likings that what customer want from a product that is sound quality. It is equally important to know that such review cannot be ignored. Most of the times, such opinions are not included in the area of opinion mining.

Moreover, opinion mining focus only on product features but product features cannot cover all significant issues in customer reviews. For example, most of the customers when talking about Nokia 6610 phone, they have written about flip phone. These reviews are critical to understand the rationale of decision making and purchase. But opinion mining doesn't consider flip phone as a feature of product.

Note: Due to the reasons mentioned above, it has been noticed that with the use of opinion mining only, it is not possible to drive all the critical information from customer reviews. So, this paper has focused on text summarization technique to summarize the product reviews including all the details discussed in customer reviews ^(Jiaming Zhan, Han Tong Loh, Ying Liu, 2009).

In previous works, opinion mining (OM) or sentiment analysis was mainly focused to extract the sentiments from customer reviews. OM mainly determines whether the reviews provided by customer's state positive, negative or neutral orientation about the product. But it is not enough to study only about positivity or negativity about reviews. In other words, concentrating only on positive or negative comments is not sufficient to cover all critical topics and different views regarding different kind of reviews. So, this paper focus on a technique that will generate a summary based on the essential tops mentioned in reviews. Thus, a summary obtained will help customers in taking a decision about buying a product or not.

In this internet age, there are many different techniques are used today to predict appropriate evaluation regarding which product to buy by concluding the overall reviews of a product given by different consumers. Such techniques include feature based summarization, summarization by fuzzy logic, summarization through lexical chains, frequent pattern mining algorithm and many more. This paper focuses on combining several techniques so as to establish a new method to generate summary of reviews in an efficient and effective way.

1.3 Different Level of Analysis

1.3.1. Document level: In this kind of methods the entire document is regarded as an only entity and methods used to do analysis are applied on the entire document. The results obtained at the document level are sometimes not correct.

1.3.2. **Sentence level:** In the sentence level methods, each and every sentence is regarded as an individual unit and approaches used to do analysis are applied on individual sentence and then the result obtained is summarized to give the overall outcome of the document. It is also called as clause level analysis

1.3.3. **Entity and Aspect level:** In this level, feature based approaches are used in which positive and negative sentiments are considered for determining the quality of features. This method is based upon the concept of opinion mining and text summarization. (Dr. Ritu Sindhu, Ravendra Ratan Singh Jandail, Rakesh Ranjan Kumar , 2014)

1.4 Text Summarization Approach

The approaches used for text summarization can be divided into two kinds which are known as extraction and abstraction. The meaning of extraction summary is that it chooses sentences or phrases from the original document having the maximum score and collects all those sentences together to generate a new as well as shorter text without varying the text written in source document. And abstraction summary is used to generate summary by using linguistic methods to observe as well as to infer the text. Other two categories can be categorized into fusion and compression. The technique named as fusion combines extracted parts consistently and compression technique focus on not to include unimportant sections of the text. (McKeown) The summarization process has three phases which are analysing the source text, defining its significant opinions, and producing the correct output.

A summary can be indicative, informative, or critical:

1.4.1. **Indicative summaries** use the standard information retrieval approach: They provide plenty of data to help users to redirect to relevant sources where they can read in more depth.

1.4.2. **Informative summaries** behave as alternatives to the source; such summaries are obtained by collecting relevant information in a precise structure.

1.4.3. **Critical summaries** include informative meaningful opinion statements on as well as bring the ideas of expertise that is not available from the source alone.

1.5 Sentiment Analysis

Today people prefer to shop online as it is easy to do and can be done in the comfort of your home. Before shopping of a particular product, customers would like to read reviews of the product so as to buy a good quality product from genuine shopping site. So, reviews play an important role in making decision. So, the World Wide Web can be viewed as the repository of reviews from customers which are spread all over the websites. Netizens look forward to the opinions to judge the product features and quality. So, due to large volume and high dependability on reviews, manufacturers and retailers face the challenge of automatically analysing the large amounts of data in form of reviews.

Today, it is very difficult to make sense out of the huge amounts of data generated every second. Using the combination of data aggregation techniques, natural language processing (NLP), linguistic analysis and proper visualization techniques, a summarized feedback would be generated in form of easy to understand graphs. This can be achieved by performing detailed sentimental analysis. Opinion mining and sentimental analysis are related to each other but a bit distinct. As opinion mining that deals with polarity detection that is positive, negative or neutral whereas sentiment analysis focuses on emotion recognition. Opinion mining is one of the steps in sentiment analysis.

Suppose, a new mobile of HTC has been launched and lots of end users are posting their reviews on it. But is there any way that manufacturer can analyse thousands of reviews collectively and then use that data for the betterment of the product? Is it possible to analyse what users are liking or disliking about that particular HTC phone?

A sentiment can be a feeling, emotion or opinion. Sentiment analysis uses NLP, statistics or machine learning methods to extract, identify or characterize the sentiment portion of the text. Sentiment analysis also deals with analyzing the writer opinion, mood, intent for a particular topic being discussed in the text and its polarity i.e. positive or negative.
(http://en.wikipedia.org/wiki/Sentiment_analysis)

Sentiment Analysis (SA) or Opinion Mining (OM) deals with the study of people's opinions, attitudes and emotions toward a particular event, individual or topics. These two

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

terms SA or OM are interchangeable which means they have same meaning. However, some researchers stated that opinion mining extracts and analyze the opinions of people about an entity whereas sentiment analysis identifies the sentiment expressed in a text and then analyze it.

Sentiment analysis includes:

1. Gathering the reviews or opinions from different portal sites, ecommerce sites, forums and social network sites.
2. Processing the data using the rules of natural language and grammar.
3. Find out what exactly was being said about (topic features) and sentiments (emotions, opinions and polarity) people are expressing.

Five main factors to consider while performing sentiment analysis ^(<http://brnrd.me/social-sentiment-sentiment-analysis/>)

- **Topics:** topic concerns with the main areas of discussion?
- **Aspects (subtopics/attributes):** Aspects concerns with what about those topics is being talked about?
- **Sentiment:** Sentiment means opinions or expressions about the content.
- **Opinion Holder:** Who is giving opinion? Are there multiple in the same content?
- **Time:** When the content was posted?

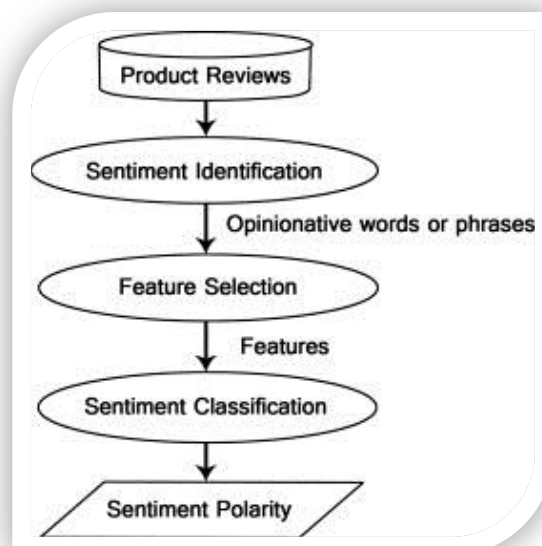


Figure1: Sentiment Analysis

1.5.1 Techniques used in sentiment analysis

a) Data aggregation techniques: The data set used for sentiment analysis includes blogs, reviews, posts, Facebook comments, tweets and many more. These posts are extracted using information retrieval techniques and combined using different aggregation rules. Data aggregation provides various tools that are beneficial in formulating collective sentiment. Collective sentiment analysis is used when sentiments are analyzed collectively for a single topic either through polarity or through scale.

b) NLP techniques: NLP techniques are used to extract the features discussed in document. Sentiment analysis uses NLP to extract the sentiments expressed in the reviews. NLP also uses in POS tagging and parsing. StanfordNLP is used for parsing the document to determine the actual sentiment being expressed.

c) Linguistic analysis: Linguistic analysis is used for deep analyzing the sentiment. Linguistic Analysis accurately handles complex structures like negation. “My new nexus is really not bad at all”. (<http://www.bitext.com/bitext-api-2/sentiment-analysis-2.html>)

The sentiment analysis should handle the complex language structures as it plays a major role in determine the actual sentiment of the sentence, such as negation or comparative sentences. Linguistic Analysis automatically handles such structures and is able to capture the difference between opinions such as:

- “This phone is much better than my old phone.” – Positive
- “This phone is not much better than my old phone.” – Negative

Sentiment scoring is also based on linguistic analysis.

d) Visualization techniques: These techniques are used to represent the result of sentiment analysis in form of graph or other representation tools. Visualization helps the users to better understand what the sentiment analysis trying to depict the overall quality of the product. For example, several visualization techniques are used to highlight different aspects of tweets and their sentiment. Like, sentiment, topics, heatmap, tagcloud, timeline, map, affinity and tweet are various tabs that present different aspects regarding sentiments.

1.5.2. Subtasks in Sentiment Analysis

1. The first task in sentiment analysis is sentiment or opinion detection which means find out whether the text is subjective or objective. Usually opinion detection is based on the examination of adjectives in sentences.
2. The second task is to determine the polarity classification i.e. to determine whether the given opinionated piece of text is positive or negative.

The above two tasks can be done at several levels including: term, phrase, sentence or at document level. Different techniques are used at different levels. Techniques like n-gram classifiers or lexicons work on term level, whereas Part-Of-Speech tagging is used for phrase and sentence analysis. Heuristics are mostly used at document level.

3. Discovery of the opinion's target. This also includes feature extraction. Given the particular topic and extracting its attributes or components of the object as its features.

1.5.3 Applications of sentiment analysis

- (i) The most common application of sentiment analysis is in the field of customer reviews regarding products. There are many online sites present which provide automated summaries of product reviews on their specific aspects. Example: "Google product search".
- (ii) Moreover, Twitter and Facebook are a focal point of many sentiment analysis applications. Sentiment analysis is used to monitor the reputation of a specific brand on Twitter/Facebook.
- (iii) Sentiment analysis is used in voting campaign. It allows campaign managers to track how voters feel about different issues and how they relate to the speeches and actions of the candidates.
- (iv) Sentiment analysis is also used in the financial markets. A sentiment analysis system use various sources to find articles which basically discuss the companies and then this system aggregate the sentiment about them as

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

a single score which can be used by an automated trading system. For an instance: The Stock Sonar.

- (v) Flyer experience tracking: A large airline company has started monitoring tweets about their flights to determine the feelings of their clients regarding delays, upgrades, new planes, services, in-flight entertainment, and many more.
- (vi) Sentiment analysis provided by Semantria offers an open-source Chrome plugin that is used to tracks author biases. This means when you run it, it will check the author name of the article which you are reading at present and then also analyze other works written by the same author. This means it pulls out his or her opinion trends by using sentiment analysis.
- (vii) Sentiment analysis is used to make brand reputation management. It is product and company rather than customer focused. Online Commerce sites should have the OpenText feature which provides both sentiment analysis and the semantic navigation capabilities. It issued to influence the purchases. (<http://smartdatacollective.com>)

1.5.4 Challenges in the sentiment analysis

1. Veracity of the data and authenticity of data: the rate at which reviews are generated on the internet is very fast and it is unsure that users providing reviews are genuine or not. So this could be achieved by collecting large amount of reviews from the trusted sites.
2. Analysis of non-standardized data: people use different ways to express their sentiments. Now a days, many shorthand's or acronyms are used so it is a challenge while processing the natural language an analysing sentiments.
3. Sheer volume of data being generated and applying the right pre-processing filters on the data for cleaning and pruning. Use of efficient code and well established transformation techniques are used to meet these challenges. (<http://www.slideshare.net/>)

CHAPTER 2

LITERATURE SURVEY

In this chapter, the author of paper has described some of the existing approaches of opinion mining and text summarization approaches. A lot of work has been done in the field of processing customer reviews. In this section numerous studies have been reviewed and some of the papers have been considered that has been taken as motivation towards the study.

Today, majority of people want to express their opinions about the products online on various online sites. This trend has raised various techniques within the context of mining customer concerns from online product reviews.

Early approaches of mining opinions from customer reviews were OPINE and SENTIWORDNET.

Monika yadav, Pradeep Mittal has proposed a paper in which they states about web mining. Web mining is the application of data mining techniques which deals with the extraction of knowledge from content, structure and usage data. Further, author has explained types of web mining which are, web usage mining, web structure mining and web content mining. (Monika yadav, Pradeep Mittal, 2013)

V. Bharanipriyal & V. Kamakshi Prasad has developed a paper in which they have discussed the concepts of web mining and mainly focused on web content mining process. Web mining is classified into 5 main subtask namely 1) Resource finding, 2) Information selection and pre-processing, 3) Generalization, 4) Analysis and 5) Visualization. Web content mining (WCM) deals with the identification of user specific data from text, image, audio or video available on the web. This is also known as web text mining. Information retrieval and natural language processing is mainly used in web content mining. The two main approaches in WCM are (1) Unstructured text mining approach for unstructured texts (free texts) and (2) Semi-Structured (html) and Structured mining (tables) approach. Tools used in WCM are Web Info Extractor, Mozenda, Screen-

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

Scraper, Web Content Extractor, and Automation Anywhere. (V. Bharanipriya & V. Kamakshi Prasad, 2011)

SANJAY MADRIA, SOURAV S BHOWMICK has proposed a paper in which they have focused on the web data mining research in context of their web warehousing project called WHOWEDA (Warehouse of Web Data). The key objective of WHOWEDA at the Centre for Advanced Information Systems in Nanyang Technological University, Singapore is basically to design and implement a web warehouse that extracts and manages important information from the web so as to provide support for strategic decision making. (SANJAY MADRIA, SOURAV S BHOWMICK)

Bakhtawar Seerat, Farouque Azam has published a paper in which they surveyed different papers and summarizes the issues and challenges of opinion mining. Challenges in opinion mining is knowledge discovery process and then issues in KDD includes an inadequate database, poor data, limitations of tools, flaws in discovery process, real world databases, absence of summary and historical data, etc. data mining web mining, types of web mining, opinion mining, model of opinion mining has been discussed in the paper. Then, challenges of opinion mining have been discussed like feature extraction, opinion orientation, identification of comparison words, etc. (Bakhtawar Seerat, Farouque Azam, 2012)

Periakaruppan Sudhakaran, Shanmugasundaram Hariharan and Joan Lu has published a paper which stated that due to increase in the number of user reviews for any product has provided a need to develop an effective system to analyze the user reviews on web. Such reviews are useful for both customer as well as manufacturers. Mining hundreds or thousands of reviews is known as opinion mining. In this paper, author has provided some directions for the future research work, focusing on the challenges and issues.

The proposed approach

Firstly, extract the user reviews from online sites. Such reviews are unstructured reviews, as they are of free form expressed by users. Now, split the reviews into sentences by considering the punctuation marks and conjunctions present in the review. Perform the pre-processing step, stop words (meaningless) and special characters are removed. Opinion mining identifies the feature words, negations, intensifiers, opinion words, neutral words in the review sentences. Then assign the score to each word. Each sentence

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

is scored and the overall scores for all the sentences in the review are calculated.

(Periakaruppan Sudhakaran, Shanmugasundaram Hariharan and Joan Lu, 2013)

Andrea Esuli and Fabrizio Sebastiani have developed a technique named as ‘sentiwordnet’ to review the quality of products. It is considered as a lexical resource in which every WordNet synset is associated to 3 numerical scores which are Obj(s), Pos(s) and Neg(s). These three scores describes how objective, positive and negative the words contained in the synset. These scores are derivative of combination of the results that are produced by a group of eight ternary classifiers having similar accuracy levels but different classification behaviour. SentiWordNet is provided with web based GUI. Subtasks are performed as determining text so-polarity, determining text pn-polarity and determining the strength of text pn-polarity. (Sebastiani, Andrea Esuli and Fabrizio)

Soo-Min Kim and Eduard Hovy have developed an approach for automatic identification of pros and cons of review sentences. Identification of subjective words, sentences and expressions is the task of subjectivity detection. Determination of positive or negative opinion of words is the task of semantic orientation classification. Mining and summarizing opinions or reviews by deriving opinionated sentences regarding features of product are the task of review classification. In this paper, an approach is used to convey that whether the review text recommends the product or not and also provides appropriate reasons for the recommendations or no recommendations for the same product which are valuable. (Hovy, Soo-Min Kim and Eduard)

Minqing Hu and Bing Liu have developed a technique which is performed in following two steps:

1. First of all, find out the product features about which various customers have given their opinions which are also known as opinion features and provide a rank for those features on the basis of their occurrences in the order they appear in the review sentences.
2. For each product feature, determine the number of customer reviews containing positive or negative opinions. The particular reviews that express those specific opinions are attached to the particular feature. This lets the looking at the reviews by various potential customers. (Liu, Minqing Hu and Bing)

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

A paper is published by Jiaming Zhan a, Han Tong Loh a, Ying Liu proposed a technique of Summarization Based on Topical Structure which includes pre-processing step (stop words removal and word stemming) and topic identification (determine the important topics discussed in the review set) and then creating the topical structure out of these topics. The technique used here is text segmentation which segments the provided text based upon the similarity of adjacent passages and also used to determine the boundary of topics. Next step is performed as candidate sentence extraction (extracting all the relevant sentences together and add them into a pool as candidate segments for the final summary). After that post-processing and final presentation is performed (redeveloping the sentences from the collected candidate sentences and view the output and final summary to users). (Jiaming Zhan, Han Tong Loh, Ying Liu, 2009)

Ladda Suanmali, Naomie Salim and Mohammed Salem Binwahlan proposed a system that includes the following key steps: 1) Firstly, read the source text or document into the system; 2) Next is to proceed with the pre-processing step, the system extracts the individual sentences or phrases of the original document. Then distinct the input document into the individual words. Then perform stop word removal and after that execute word stemming; 3) Now, associate each sentence with the vector of eight features and the values for these features are calculated from the content of the sentences; 4) These features are calculated to determine the score of the sentence which would be based upon general statistic method and fuzzy logic method; 5) Extract the sentences having highest score and deliver as a document summary depending upon the compression rate. (Ladda Suanmali, Naomie Salim, Mohammed Salem Binwahlan, 2009)

Seyed Hamid Ghorashi, Roliana Ibrahim, Shirin Noekhah and Niloufar Salehi Dastjerdi proposed a technique that includes the following steps: 1) Pre-processing step includes two major things that are stop words removal as well as word stemming; 2) Performs POS tagging which includes identification of frequent features; 3) Mining frequent patterns that lead to determining the potential features; 4) Pruning involves compactness pruning as well as redundancy pruning and therefore frequent features will be extracted. Finally summary can be made comprising the sentences including potential features. This paper proposes a technique which includes different steps to obtain a summary from review sentences. (Seyed Hamid Ghorashi, Roliana Ibrahim, Shirin Noekhah and Niloufar Salehi Dastjerdi, 2012)

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

Pooja Kherwa, Arijit Sachdeva, Dhruv Mahajan, Nishitha Pande, Prashant Kumar Singh has developed an approach discussed in the paper that scans every line of the data and generates a logical summary of every review that undergoes a categorization by aspects and presents the result in form of graphical representation. The proposed system breaks reviews into sentences and then sentences further break down to tokens and then analyze the sentiment corresponding to each relevant token individually and then calculates overall sentiment as well. The summary doesn't output the sentences as a whole but directly the relevant features pointed out with corresponding sentiment. (Pooja Kherwa, Arijit Sachdeva, Dhruv Mahajan, Nishitha Pande, Prashant Kumar Singh, 2014)

Y. Surendranadha Reddy, Dr. A.P. Siva Kumar has proposed an approach which has been explained through paper. The approach discussed in paper provides a summary of the web document based on the sentence importance measures. A sentence is ranked on the basis of term frequency and sentence similarity measures. Rank is provided to the sentence according to the sentence score and then sentences with higher rank are selected to be included in summary. The proposed method includes

Calculate the similarity between each pair of sentences.

Calculate the frequency of the terms in each of the sentences.

Compute the score for the sentences in the specified document.

Assign ranks to the sentences according to their score. (Y. Surendranadha Reddy, Dr. A.P. Siva Kumar, , 2012)

Dongjoo Lee, Ok-Ran Jeong, Sang-goo Lee has proposed a paper in which various techniques used in multiple steps of opinion mining are surveyed. Opinion mining can be divided into three major tasks namely development of linguistic resources, sentiment classification, and opinion summarization. There are four major approaches in developing linguistic resources for OM: the conjunction method, the point wise mutual information (PMI) method, the WordNet exploring method, and the gloss classification method. Sentiment classification is the process of identifying the sentiment of the document i.e. its polarity. Three methods used for this purpose are PMI method, machine learning methods and combined NLP. Opinion summarization is the process in which product features are extracted and then sentiment of each feature is assigned. Then these are summarized and presented in various forms. (Dongjoo Lee, Ok-Ran Jeong, Sang-goo Lee)

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

Arti Buche, Dr. M. B. Chandak, Akshay Zadgaonkar has published a paper in which various techniques that have been developed for the key tasks of opinion mining has been discussed. In this paper, various data sources are considered like blogs, review sites and microblogging like Facebook, twitter, Tumbler. Then a typical approach for sentiment classification (identifying the whole document as positive or negative) is to use machine learning algorithms has been stated in paper. The news articles and web pages were automatically catalogued by using text classification algorithms which includes Naïve Bayes classifier. Red Opal is a tool which allows users to find out the products on the basis of features. ^(Arti Buche, Dr. M. B. Chandak, Akshay Zadgaonkar, 2013)

Wesley T. Chuang, Jihoon Yang has published a paper in which an automatic text summarizer is established that I used to provide summary. In this paper, firstly sentences are broken into segments with the help of special cue markers and then each segment is represented by a set of predefined features like location of the segment, term frequencies of the words occurring in the segment, number of title words in the segment and many more. Then a supervised learning algorithm is used to train the summarizer that will extract the important sentence segments. Sentence segmentation: Sentences are divided into segments through cue words like because. Feature Representation: There are two kinds of features we consider: structured and non-structured. The former are related to the structure of the text (e.g. rhetorical relations), while the latter are not (e.g. title words). In order to train summarizer, decision trees or Naïve Bayes classifier are used. ^(Wesley T. Chuang, Jihoon Yang,)

G.Angulakshmi, Dr.R.ManickaChezian has published a paper which discusses about opinion Mining and its techniques and tools used. Data source used to collect the data for opinion mining includes blogs, review sites and data set. Opinion mining and sentiment analysis been discussed in the paper which talks about following terms:

Direct opinion (deals with positive or negative directly) or comparison (comparing two objects), Pre-processing (tokenization, stop words removal, case normalization and word stemming), Feature extraction deals with feature types, feature selection (good features are selected for opinion classification), feature weighting mechanism, reduction mechanisms.

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

Feature types deals with term frequency (number of time features occurs), term co-occurrence (feature occurring together- unigram, bigram), part of speech information (POS tagger), opinion words, negations, syntactic dependency.

Feature selection includes information gain (based on the presence and absence of a term in a document and the term having less than threshold is removed) Odd Ratio (term having one positive and one negative class for classification, Document Frequency measures the number of appearances of a term in the available number of documents.

Features weighting mechanism could be of type 1 i.e. Term Presence and Term Frequency or could be of type 2 i.e. Term frequency and inverse document frequency (TFIDF) (G.Angulakshmi, Dr.R.ManickaChezian, 2014)

Walter Kasper, Mihaela Vela has developed a system for the hotel for better hotel management that collects comments from the web and creates classified and structured overviews of such comments and provides facility to users to access that information. The system present in the paper is a part of BESAHOT. People who want to get actual overviews about hotels on the web can take advantage of such system. The core system on the server side handles data acquisition, analysis and storage. The acquisition of reviews coming from the web is handled by a web crawler which includes HTML pages as well as RSS feeds. Filter patterns are used which restrict the crawler to crawl only relevant links. After retrieving the target page, content extraction module is applied that will extracts the relevant textual content of the review and other metadata like information about the reviewer. Analysis part undergoes the segmentation of the review which further undergoes static polarity classifier and linguistic analysis (tokenization, POS tagging, stemming, spellcheck, information extraction). The results of both go to joint polarity. (Walter Kasper, Mihaela Vela)

Raymond Kosala, Hendrik Blockeel has proposed a paper which explains all about web mining. Web mining is the application of data mining techniques to extract knowledge from web data. It consists of three parts namely, content mining which is the process of extracting useful information from the contents of web documents. Issues addressed in text mining are topic discovery, extracting association patterns, clustering of web documents and classification of web pages. Research activities of this topic have drawn heavily on techniques developed such as Information Retrieval (IR) and Natural Language Processing (NLP). Next is web structure mining which is a typical web graph

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

consists of web pages as nodes and hyperlinks as edges connecting related pages. Web structure mining is the process of discovering structure information from the web. This can be further divided into two kinds which are hyperlink and document structure. Then, web usage mining is used to discover interesting usage patterns from web data to understand and better serve the needs of web based applications. (Raymond Kosala, Hendrik Blockeel)

Freimut Bodendorf, Carolin Kaiser has developed a new approach which is based upon text mining and social network analysis which helps in detecting opinion leaders and opinion trends. A new approach starts with detecting opinions and relationships among forum users by text mining. With the help of social network analysis, opinion leaders are identified and opinion evolution is analyzed. The analysis consists of four steps. Firstly, opinions of users on the specific product are extracted. Then the communication relationships among users are identified by text based relationship mining methods. Now, represent the extracted users, opinions, and relationships form a social network in form of graph. The nodes represent the users of a forum and the edges their communication relationships. The resulting graph is analyzed by determining key figures for the position of single users and for the overall structure of the network. In this way opinion leaders and opinion trends can be identified. (Freimut Bodendorf, Carolin Kaiser)

Dr. Ritu Sindhu, Ravendra Ratan Singh Jandail, Rakesh Ranjan Kumar have proposed a paper which uses clustering and classifying technique for opinion mining by analysing the blog post on recent and services reviews. Different level of analysis (document, sentence, entity/ aspect level) is being discussed in paper. Next thing discussed in paper is about issues and applications in sentiment analysis.

Methodology

Firstly, split the document into sentences and then every sentence according to its problem domain, if sentence is not belongs to any domain then it will be consider as the normal sentence. Now assign the weight to each sentence like number of available keywords will be the weight of that particular sentence. In case of product, the product will be classified in its attributes and on the basis of keyword available for a particular attributes assign a weight ranging from -5 to +5. In this way summaries about the product can be found and also the individual attributes available in the product can be determined.

(Dr. Ritu Sindhu, 2014)

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

Minqing Hu and Bing Liu have developed an approach which performs summarization task which is different from traditional text summarization as authors are only interested in the specific features of the product that customers have opinions on and also determines whether the opinions are positive or negative. In this paper, we only focus on mining opinion/product features that the reviewers have commented on. There are plenty of techniques which are used to mine such features named as opinion summarization system.

Summarization is done in two main steps that is feature extraction and opinion direction identification. POS tagging: In this NLP linguistic parser is used to parse each sentence that yields the part-of-speech tag of each word (determine whether the word is a noun, verb, adjective) and identifies simple noun and verb groups (syntactic chunking). Frequent Features Generation (Apriori can be used) is used find features that people are most interested in. Feature pruning includes compactness pruning and redundancy pruning. Then at last opinion words are extracted and opinion orientation is determined.

(Liu, Minqing Hu and Bing)

R.V.V Murali Krishna and Ch. Satyananda Reddy have published a paper which established a new methodology for implementing the stoplist concept and statistical analysis concept based upon parts of speech tagging. A sentence scoring mechanism has been developed by combining the above methodology with semantic analysis.

The proposed system

A sentence scoring method is built using the concepts of stop word removal, Semantic relationship and statistical relationship.

Algo1: Algorithm for Stop Word Removal

Step 1: Parse the sentence in to words based on Standard English language tokens.

Step 2: Tag the words with their corresponding parts of Speech.

Step 3: Add the words to the keywords list whose tag is Noun/verb/adverb/adjective.

Similarly, algorithms for sentence processing, semantic analysing, statistical analysing and weighted average have been discussed in paper. (R.V.V Murali Krishna and Ch. Satyananda Reddy,

2012)

Ali Harb, Gerard Dray, Michel Plantié have developed an approach for extracting the opinions from blogs focusing on two steps. Firstly, automatically extract a learning dataset for a specific domain. Then, extract, the set of positive and negative adjectives

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

relevant for the domain from this learning set. This approach is known as AMOD approach which consists of 3 phases: Corpora Acquisition learning phase, Adjective extraction phase and Classification. (Ali Harb, Gerard Dray, Michel Plantié)

Subhabrata Mukherjee, Pushpak Bhattacharyya presents a paper in which an approach is defined which identifies feature specific expressions of opinions present in product reviews containing different features and mixed emotions. This approach deals with identifying a set of potential features in the review and then extracting the opinion expressions about those features by exploiting their associations. Dependency parsing is used to identify relations between the opinion expressions. (Subhabrata Mukherjee, Pushpak Bhattacharyya)

Ana-Maria Popescu and Oren Etzioni have introduced a system Opine which is used to mine the reviews. OPINE is a high-precision opinion mining system which extracts fine-grained features and associated opinions from reviews. OPINE successfully uses the Web in order to improve precision. It extracts positive/negative opinion words. Then classify reviews as positive or negative and then identify feature-opinion pairs together with the polarity of each opinion. (Ana-Maria Popescu and Oren Etzioni)

MR. S. M. VOHRA, PROF. J. B. TERAIYA states that Sentiment analysis is the automated mining of attitudes, opinions, and emotions from text, speech, and database sources by using Natural Language Processing (NLP). Analysis and comparison is done between machine learning techniques, lexicon based techniques and hybrid techniques. (MR. S. M. VOHRA, PROF. J. B. TERAIYA)

C. Lakshmi Devasenal and M. Hemalatha have proposed a paper in which a text analyzer is developed which is used to derive the structure of the input text using rule reduction technique in three stages namely, Token Creation, Feature Identification and Categorization and Summarization. This Automatic Text categorization and summarization is the process which automatically assigns pre-defined class labels to incoming, unclassified documents. Firstly, create the tokens for the given input, then recognize the features of the created tokens and then finally categorize the alpha tokens and summarize it into sentence. (C. Lakshmi Devasenal and M. Hemalatha, 2012)

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

Alok Ranjan Pal, Diganta Saha has developed an approach that performs the summarization task using unsupervised learning methodology. The importance of a sentence occurring in an input text is evaluated by using Simplified Lesk algorithm. By using online semantic dictionary WordNet, first of all evaluates the weights of all the sentences of a text separately using the Simplified Lesk algorithm and then arranges them in decreasing order according to their weights. Next, according to the given percentage of summarization, a particular number of sentences are selected from that ordered list. ^(Alok Ranjan Pal, Diganta Saha)

E.PadmaLahari, D.V.N.Siva Kumar, S. Shiva Prasad have proposed a technique using both linguistic and statistical features using successive threshold in order to find out the effective summary including important sentences from the given input text document. In this technique, sentences are selected for summary according to the weight of and weight is provided using both linguistic and statistical features. Sentences are given scores by weighting the features like term frequency, word occurrences, and noun weight, phrases etc. Statistical features includes: Keyword Features, Sentence Position, Term frequency, Length of the Word, Parts of speech tag whereas linguistic features include Proper Noun feature and Pronouns. ^(E.PadmaLahari, 2014)

MAQBOOL AL-MAIMANI, NAOMIE SALIM, AHMED M. AL-NAAMANY has proposed a paper which presents a review covering the semantic and Fuzzy-based logic techniques in sentiment analysis. This paper has discussed about advantages of feature-based approach. Then techniques for the semantic aspects are discussed which includes NLP covering POS, tokenizing, N-Gram, term presence, stemming, and lexicon-based methods using lexical based dictionaries like WordNet. Machine learning which can classified as supervised learning like SVM and BN classifiers. Unsupervised ones like PMI-IR classifier, Rule-based techniques, Statistical techniques, semantic using semantic web techniques many more. ^(MAQBOOL AL-MAIMANI, NAOMIE SALIM, AHMED M. AL-NAAMANY, 2014)

Amandeep Kaur and Vishal Gupta have established a paper in which various techniques used in sentiment analysis and opinion mining have been discussed. Popular approaches used for sentiment analysis includes Subjective lexicon (containing list of words where each word is assigned a score that indicates nature of word in terms of positive, negative or objective).

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

N-Gram modelling (uni-gram, bigram, tri-gram or combination of these) is used for classification. Machine learning (performs the supervised or semi- supervised learning by extracting the features from the text. (Amandeep Kaur, Vishal Gupta, 2013)

YuanbinWu, Qi Zhang, Xuanjing Huang, LideWu have proposed a paper in which authors have given a novel approach for mining opinions from product reviews. In given approach, opinion mining task is converted to identify product features, expressions of opinions as well as relations among them. By observing that a lot of product features are regarded as phrases, then a concept of phrase dependency parsing is introduced, which the extension of traditional dependency to phrase level. This concept is then used for extracting relations between product features and expressions of opinions. (YuanbinWu, Qi Zhang, Xuanjing Huang, LideWu.)

Pravesh Kumar Singh, Mohd Shahid Husain has made an attempt to review the various techniques used for opinion mining as well as sentiment analysis. Classification techniques used for opinion mining includes naïve bayes classifier, support vector machine, multi-layer perceptron, clustering classifier and many more. (Pravesh Kumar Singh, Mohd Shahid Husain)

CHAPTER 3

PRESENT WORK

3.1 Problem Formulation

The existing model is based upon the product review analysis on the basis of sentiment analysis of the user reviews. The existing system evaluates the positivity and negativity of the user reviews and returns the rating of the product being reviewed. The existing model classifies the messages as positive or negative messages. The existing system does not classify the user reviews in appropriate class or product class or product group. Also, the existing system does not deeply analyze the sentiment in terms of further positive or negative emotions like joy, cheer, satisfied, happy, sad, unsatisfied, rage, etc. These emotions may give much better product review than the existing model. The existing model is not capable of extracting or learning the new categories from the user reviews. By automatically learning the categories, the system can add more types of categories of various titles of similar category in order to evaluate the user reviews with the higher accuracy using the highly classified data.

3.2 Objectives of the study

The primary question for text summarization regarding customer reviews is to predict the quality of product. Customers want to know whether to buy product or not by reading summary of reviews generated by a text summarization approach. The problem is that most of the existing methods focus on opinion mining concept when processing the reviews provided by customers, which is then used to determine the attitude of reviewer's which can be either positive or negative regarding various certain features of product. But there are some reviews which cannot be stated as either positive or negative but these are valuable reviews and cannot be ignored. So, for encountering such reviews various methods are used. One of the appropriate approaches is to provide a summary of the product quality which could help customers to take decision accordingly.

Another major problem is that there are hundreds or thousands of reviews are present online and it is not possible to reach each and every review in order to make a buy or leave decision. In order to solve such problems, summarization approaches are widely used to determine the overall quality of product.

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

The difficult part is that there could be mixed opinions regarding product in a document, and people may convey the same opinion in vastly diverse ways. Hence, there are several techniques used in generating summary of hundreds or thousands of reviews in order to predict whether the product is good or not. Several methods are using several different approaches. The author of the paper has decided to develop a new technique by combining several existing techniques to produce summary in an efficient and effective way.

The **main objective** is to develop an optimized summarization approach that provides an effective summary of product on the basis of customer reviews so that it would help the interested customers to take the decision whether to buy the product or not to buy.

The research is focused on following objectives:

1. To help the customers in taking right decisions regarding purchase of the product.
2. To provide an effective summary of the product reviews to the customers who are interested in buying that particular product.
3. To analyze the different aspects of product i.e. what is negative about the product? What are the features that attract the customers to purchase?
4. To develop an optimized approach of providing the summary of the product on basis of customer reviews.
5. To provide the overall sentiments about the product in terms of its features according to the reviews provided by consumers. Pictorial representation is done for showing the product opinions in terms of its features.
6. To provide the comparison of the two different models of a product and shown which model is better?
7. To provide a system that asks for user requirements and then providing the best product according to the user requirements.
8. To provide the summary that includes about every product feature.

In order to achieve the objective of the method, this approach require the customers to select what kind of review, he or she is going to enter; the subject of the review and then detailed review of the product. This approach uses the pattern matching techniques to

match the subject of the review with detailed review. Many different factors like review length, sentence position and many more, are considered for review scoring as well as sentence scoring in order to provide high rank to good reviews as well as useful sentences within the good reviews.

The Proposed System

The proposed model will be based upon the deep sentiment analysis of the user review on the social networks or on the company or product websites. The proposed model will be capable of evaluating the emotions in the form of positive and negative emotions on the basis of keyword score. The keyword extraction or parsing would be done for single and multiple keywords in the various stages of the emotion analysis process. In the proposed model, the emotion analysis will be performed using the various emotions like joy, cheer, satisfied, happy, sad, unsatisfied, rage, etc. which will give us higher level of analytical product review by user reviews, the product categories will be learnt by using a product learning method. The automatic classification and learning will give a deep product review analysis.

3.3 Research Methodology

The author has proposed the method to produce the summary in an optimized way. The proposed method consists of major three steps namely, Tokenization, Keyword Matching and Summarization. The above three steps are briefly discussed as follows:

1. **Tokenization:** It is the process of dividing the whole document into paragraphs and then paragraphs into sentences. Furthermore, sentences are further divided into words. Tokenization process also includes filtration process. Filtration means stop word removal and word stemming. The result obtained after tokenization consist of tokens also known as keywords which can be noun, adjective, verb, etc. These are the sort of words which clearly conveys the opinions of customers either positive or negative.
2. **Keyword Matching:** Now, match the tokens with the knowledge-data in order to extract the message score. This step is somewhat similar to the SentiWordNet. For example, look out for any token whose score need to be calculated from the SentiWordNet text file. This step is explained by an example,

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

We need to calculate: Score (Awesome).

- a) Look up “Awesome” in the SentiWordNet text file as an adjective.
- b) It was found. So, the Score (Awesome) = 0.75.

Another example: Let’s assume we found a certain dependency as: phone = [love]

Calculate: Score (love)

- a) Look up “Love” in the SentiWordNet text file as an adjective.
- b) It returned null. So, search for “Love” as a verb again.
- c) It was found. Now, since faint verbs like “fight”, “run”, “walk” etc. do not describe a noun as effectively as an adjective, we need to filter these by passing them through a certain threshold, say 0.4.

Since Score (love) = 0.45971, the verb is strong enough to be considered influential.

So, Score (love) = 0.45971.

Classify the message according to the score as well as product feature.

3. Summarization: The output will be shown in form of bar graphs. The author has taken cell phones under study in which certain models of phone are compared with respect to some parameters (say, touch screen, camera and sound quality). After the graphic display, summarization will be produced on the basis of the visual information and classification information array.

Sentiment Analysis is the technique to analyze the emotion from a text document, message, or similar content. Sentiment Analysis algorithms are used to predict the public opinion on various issues getting discussed in the social network threads. Hence, it is also called opinion mining or emotion mining. In this paper, the author has proposed the algorithm to calculate the sentiment in the social threads with positive or negative messages. At first, the messages are scanned for their positivity or negativity. Then, the messages are automatically classified according to their product feature and re-analyzed

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

for further emotion. The algorithm uses the weighted dictionary matching to analyze the sentiment in the messages. The social thread is also recording the results on the basis of every user taking part in the social discussion thread. The results have shown the high accuracy of 97.3%. The accuracy is measured on the basis of type 1 and type 2 statistical errors.

The author has proposed the usage of a rich set of sentiment analysis features like positive, negative, automatic product feature classification and automatic summarization. The proposed feature selection method can improve opinion classification performance. The proposed Feature Relation Network is rule-based sentiment classification method that finds the text features and emotions from the given message data. The proposed algorithm consisted of five basic components: Post/Thread Acquisition, Tokenization, Polarization, Negative Emotion Analysis & Comparison module.

Post/Thread Acquisition: The first step is used to read the post saved in the excel file. The comments are read and classified as the user comments by grouping the comments of users in one group.

Tokenization: In this module, the program reads the known words from the user comment based on the word list prepared on the basis of most common words. This word list is obtained by matching the words in the user comments along with a valid word list file already saved as a text file. The text file is loaded into the memory and passed to the tokenization process for the further computations. The process of tokenization then extract all of the words from the user comment and filter them on the basis of a list of common words containing no emotion. The filtered list is obtained after deleting those matching words in the user comment. These common words are not given in the word weight file containing the rank/weight of each word being used in the common English language, which contains a neutral, positive or negative emotion.

Polarization (Positive or Negative rating module): The user comments are polarized in three major categories under this step. The three major categories are positive, negative and neutral. The tokenized comments are compared with a list of words. The file contains the ranking for each of the word listed on the list. The rank or weight or strength of the words has been listed in the document, which ranges between -5 to +5. The words are

classified on the basis of their use and its impact in the natural English language spoken in our daily lives.

Negative Emotion Analysis: All the user comments marked as negative then undergoes the negativity analysis, which checks the comments for the different negative emotions. The user comment is compared with two different files, out of which one is containing the words representing product feature classification and other for automatic summarization. The comment is marked on the basis of higher weight. For example if public review on the specific product is found high positive then the text summarization module shows the significance emphasis on the emotion analysis.

Comparison between models: The reports of individual cell phone models which are selected by user for comparison are being compared on the basis of their features.

Algorithm 1: Brief Design of **P**roduct **R**Eview and **S**UMmarization (**PRESUM**)
Algorithm

1. Obtain the data from the social network thread Tr
2. Extract the list of users U from the social networking thread
3. Extract N number of words from Message M using dictionary based tokenization
4. Filter message content with STOPWORD list of common English words while Tokenization
5. Load product and feature classification (PFC) knowledge data
6. Classify the message after comparing it with the PFC data
7. Acquisition of the sentiment and expression classification (SEC) knowledge data
8. Calculate the message score after comparing it with SEC data
9. Classify the message according to the score and increment the product or product feature index accordingly
 - a. If score is more than zero
 - i. Increment the positive index
 - b. If score equals zero
 - i. Increment the neutral index
 - c. If score is less than zero
 - i. Increment the negative index

10. Load product review summarization (PRS) knowledge data
11. Prepare the summarization content according to the sentiment report

Algorithm 2: Detailed Explanation of Sentiment Analysis Model

12. Obtain the data from the social network thread **Tr**
13. Extract the list of users **U** from the social networking thread
14. Extract **N** number of Message **M** using dictionary based tokenization
15. Filter message content with **STOPWORD** list of common English words while Tokenization
16. Load negative and positive sentiment expression word classification information file
17. Calculate word weight score to measure the sentiment **Sn**
18. Count the final sentiment score **S** of each message (Positive/Negative)
19. Calculate sentiment score for tokenized message number **N**
20. Find the sentiment type **St** (Positive/Negative) by validating the sentiment specific word dictionary
21. If **St >0**
 - a. Mark the message as positive
 - b. Add 1 to **posMsg**
22. If **St ≤ 0**
 - a. Mark the message as negative
 - i. Add 1 to **negMsg**

3.1 WORKFLOW DIAGRAM

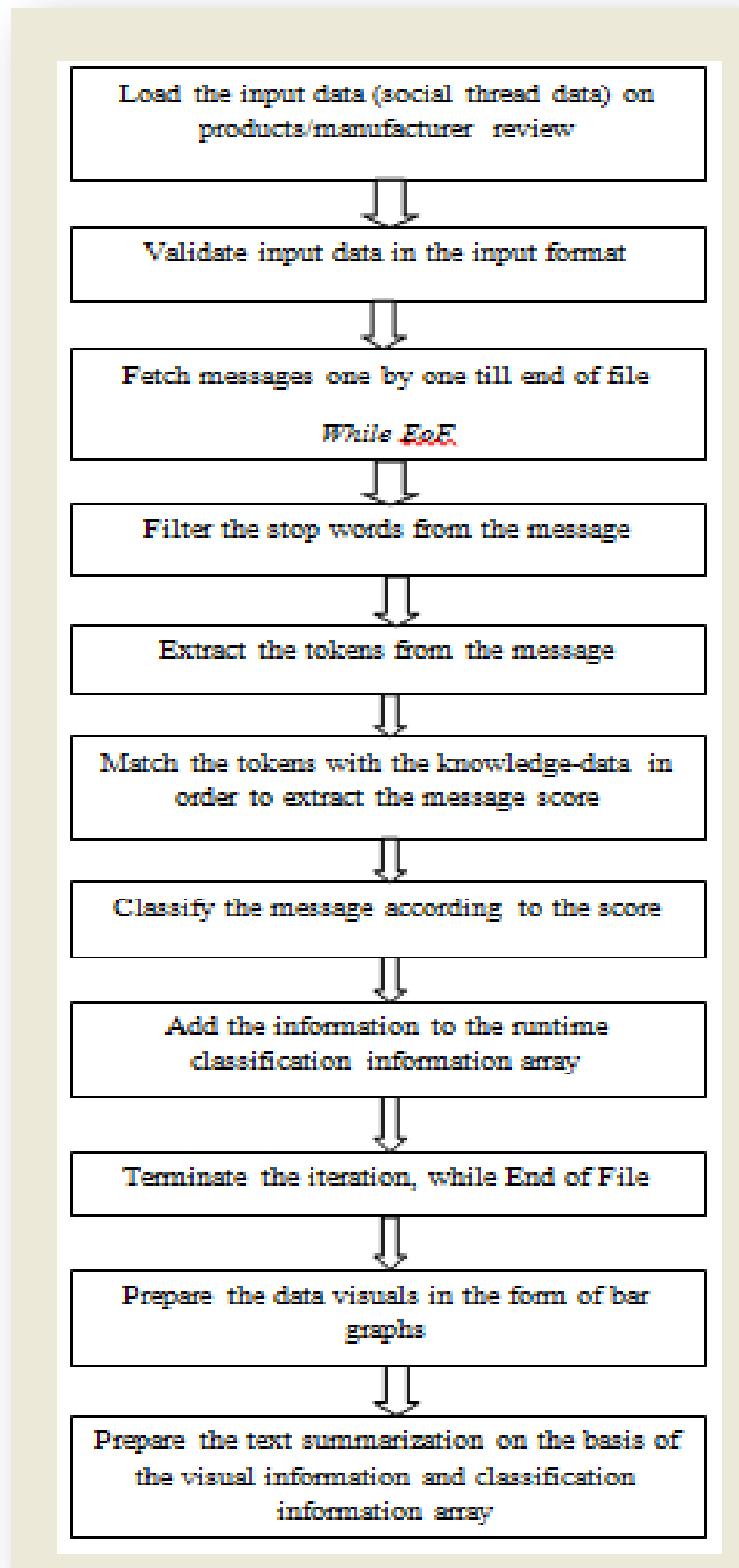


Figure2: Flow Diagram

CHAPTER 4

RESULTS AND DISCUSSIONS

The results have been obtained from the proposed model. The proposed model have been given a discussion thread from the social network website collected in the excel sheet as the input data. The sentiment analysis has been performed on the excel file containing messages to find certain emotions automatically. The proposed algorithm returns the positive, negative, automatic product feature classification after analyzing the messages. The emotions are calculated by analyzing the words weight in the certain combinations & counting the whole emotion weight to calculate the resultant weight of the message. The messages are firstly broken in the words, phrases or combination of words, collectively called tokens, matching with the pre-programmed dictionary file stored up in the proposed model.

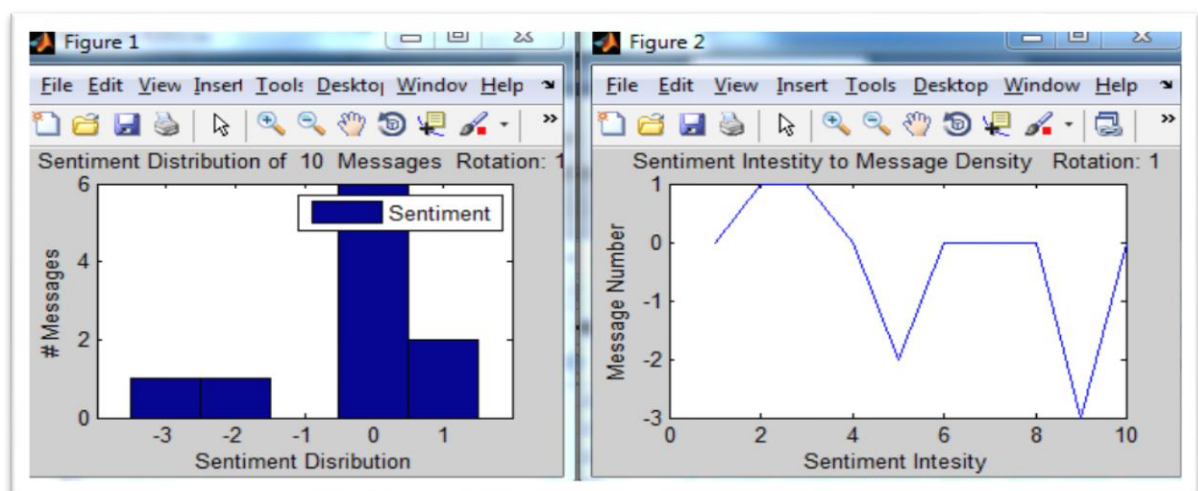


Figure3: Sentiment Analysis on 10 Messages

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

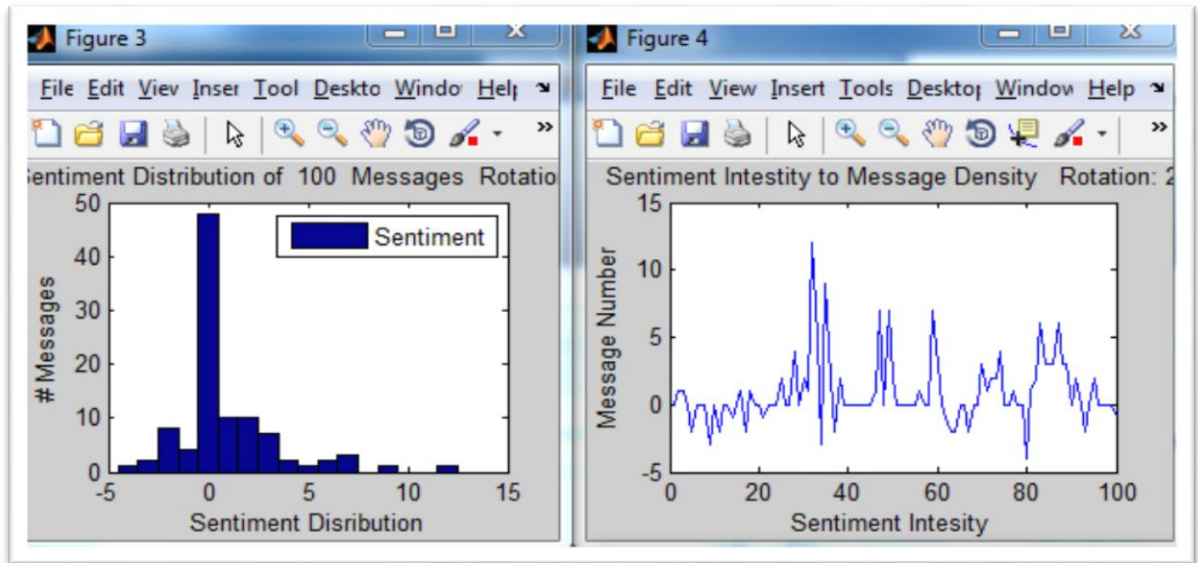


Figure4: Sentiment Analysis on 100 Messages

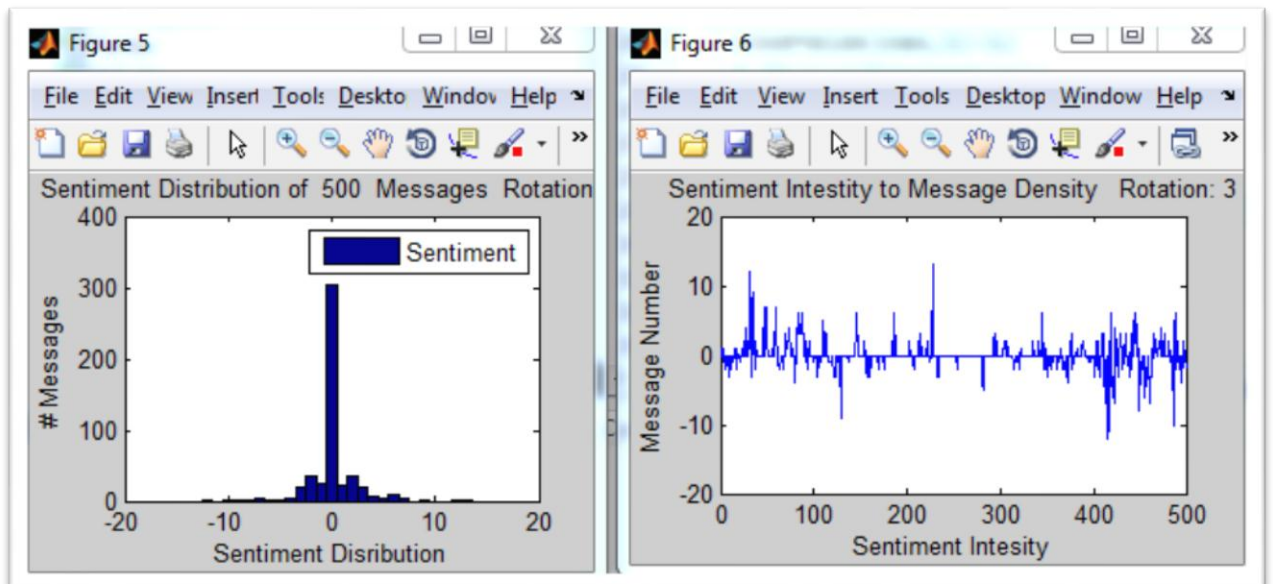


Figure5: Sentiment Analysis on 500 Messages

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

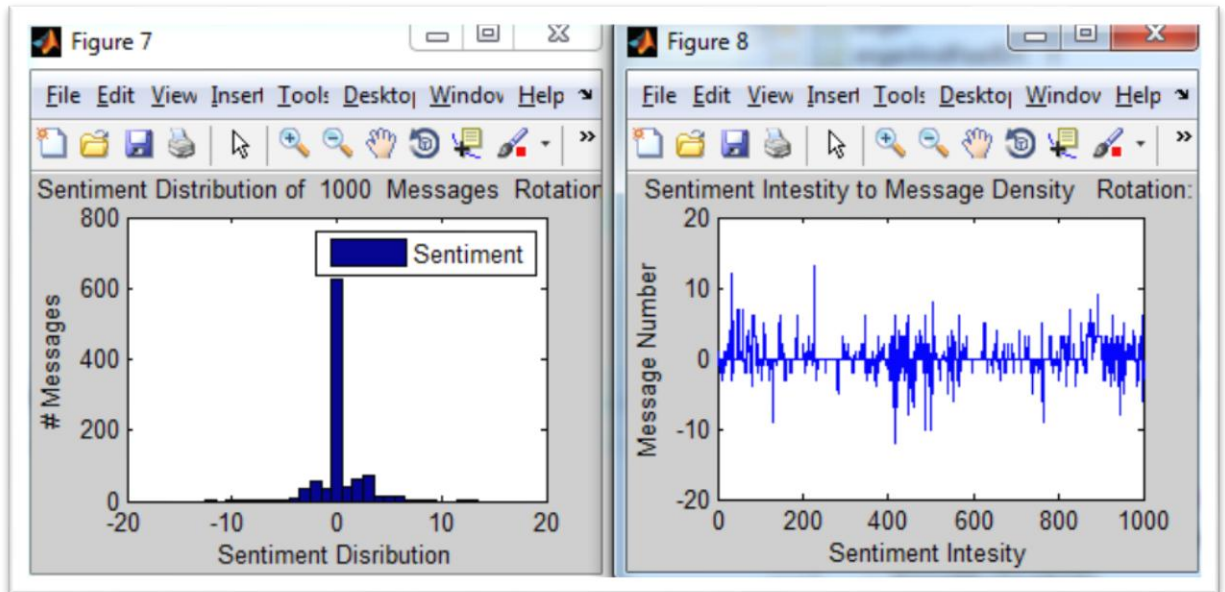


Figure6: Sentiment Analysis on 1000 Messages

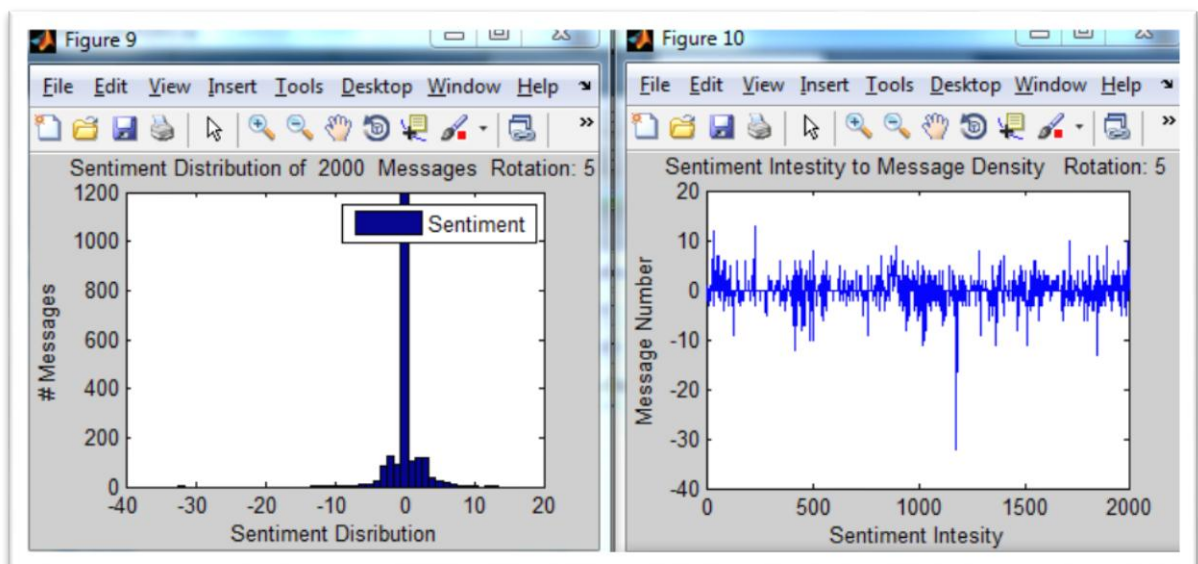


Figure7: Sentiment Analysis on 2000 Messages

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

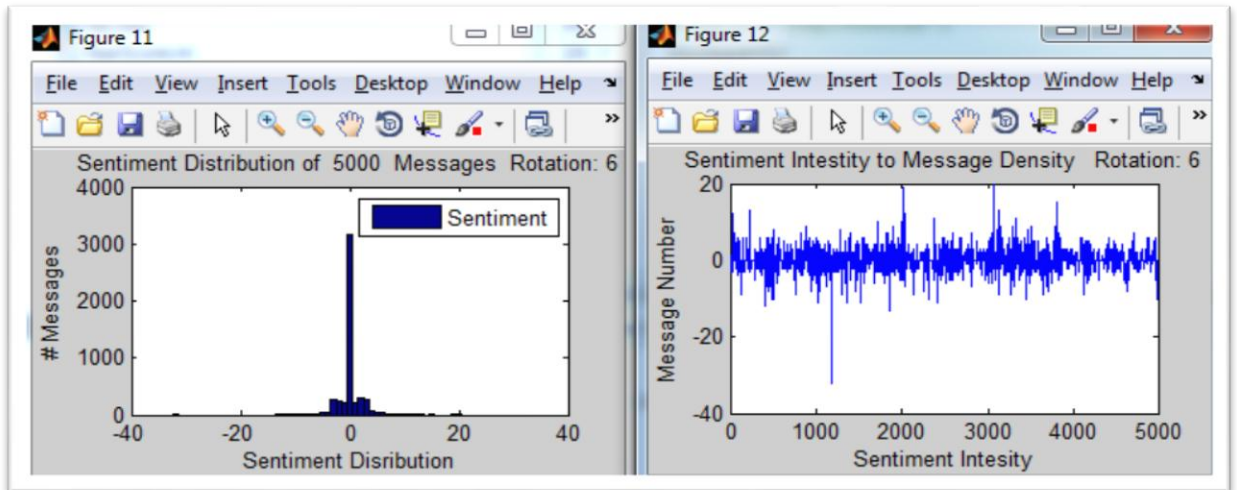


Figure8: Sentiment Analysis on 5000 Messages

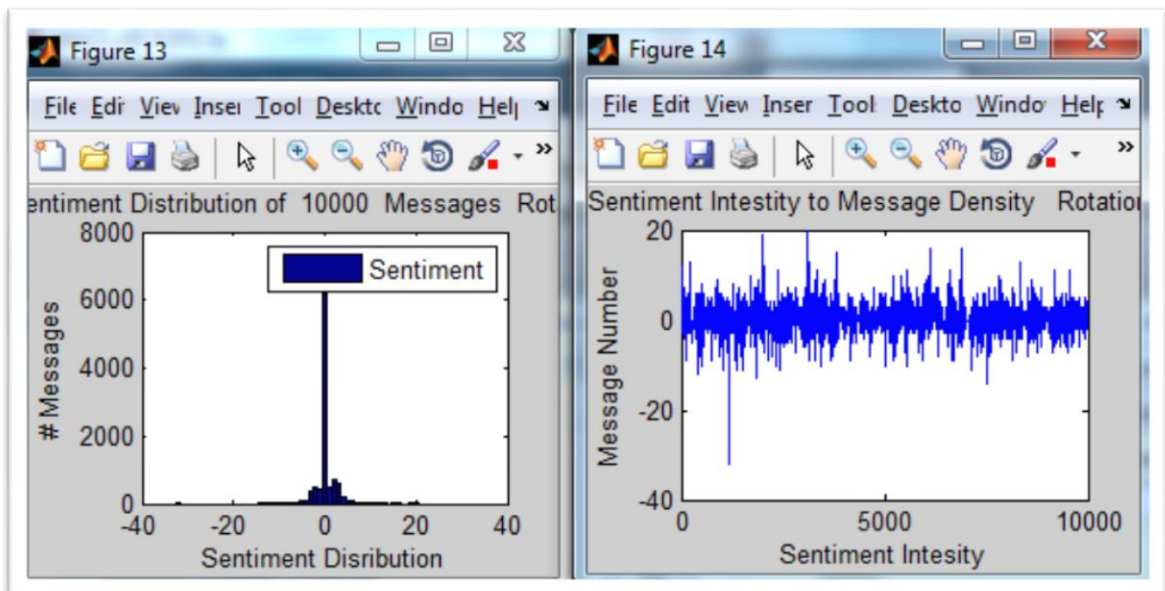


Figure9: Sentiment Analysis on 10000 Messages

Then the score or weight is calculated for each token, & a total of the token score is calculated and published. The published score is analyzed against the slab to mark the message as positive or negative. The dataset messages are analyzed against the pre-programmed dictionaries for positive and negative words and their scores. After the calculation of the emotion score, the message is marked as negative or positive and further classified according the product feature, where the message is again analyzed using the sentiment analysis module which gives the user review on that specific feature. The number of negative messages, positive messages has been given in the table 1. Table 1 is also equipped with the number unreadable messages from the other languages. Table 1 shows the normal sentiment analysis using the proposed scheme. The following table

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

data has been designed to measure the accuracy of the proposed model in terms of analyzing the emotion from the given dataset. The accuracy has been tested against the manual classification. The manual classification is entirely based upon the natural emotion selection by the human.

Table 1: Analysis of various numbers of messages

Total Number of Messages	Positive Messages	Negative Messages	Message From Other Languages
10	2	8	0
100	37	63	5
500	100	400	24
1000	219	781	24
2000	430	1570	27
5000	984	4016	47
10000	2190	8810	212

Table 2: A Table of Statistical Error Calculated manually on the given Database

Total Number of Messages	True Positive	True Negative	False Positive	False Negative	Recall (TP/TP+FN)	Precision TP/TP+FP
10	10	0	0	0	100%	100%
100	94	1	5	0	100%	94.90%
500	483	5	12	0	100%	97.57%
1000	978	8	14	0	100%	98.58%
2000	1943	24	29	2	99.80%	98.52%
5000	4897	43	56	3	99.91%	98.86%
10000	9790	89	113	8	99.18%	98.85%

The results have been obtained by applying the proposed algorithm for 10, 100, 500, 1000, 2000, 5000 and 10000 messages. The overall percentage of correct results by the proposed algorithm is 97.3. The algorithm has shown a higher accuracy in term of sentiment analysis.

In this thesis project, the proposed model has shown significant accuracy in calculating the sentiment in the social thread. The social thread collected from social network Facebook, Twitter or GSMarena has been analyzed with this proposed algorithm for analyzing its performance. The results in the table 1 are showing the different types of emotions calculated on the messages using the proposed algorithm. The accounted emotions the message been marked with are positive, negative, product feature classification and unreadable messages from other languages. The proposed algorithm has been tested with various sizes of messages data in each rotation as given in the table 1. The results of type 1 and type 2 statistical errors given in table 2 are manually validated.

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

The results are proving the good performance of the proposed algorithm with accuracy of almost 97%.

In the future, the proposed algorithm can be enhanced to calculate product features of more products or with a wider range. The above emotions can be probably calculated using the dictionary based phrase specification methods. Also, the proposed algorithm can be improved on the basis of execution time and accuracy.

MOBILE PHONE REVIEW RESULTS

The proposed model has been designed to analyze the public opinion the mobile phones. The public reviews on the handsets from the various manufacturers have been obtained from online sources such as Facebook, Twitter and GSMarena. The public reviews has been kept in the excel format and the loaded in the runtime memory using the MATLAB module for Microsoft document handling. The user review data has been automatically classified for the product features and then the emotion analysis has been performed on the classified messages and the calculation emotion is added to the corresponding category under the product review classification. The product review classification has been done using the dictionary defined for the various product features in the mobile phones. The result snapshots have been shown in the following section:

RESULT COMPARISON WITH EXISTING MODEL

Table 3: A Table of Precision and Recall for various Densities of Messages

Sr. No.	Total Number of Messages	Recall (TP/TP+FN)	Precision TP/TP+FP
1	10	100%	100%
2	100	100%	94.90%
3	500	100%	97.57%
4	1000	100%	98.58%
5	2000	99.80%	98.52%
6	5000	99.91%	98.86%
7	10000	99.18%	98.85%
AVERAGE	-	99.84%	98.18%

The proposed model has been evaluated for the various performance parameters which show its effectiveness using the Recall and Precision values. The Precision and Recall of the proposed model is quite higher making the proposed model highly efficient in the terms of sentiment analysis, product feature classification and automatic summarization.

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

Table 4: The Table of Comparison between Existing and Proposed model

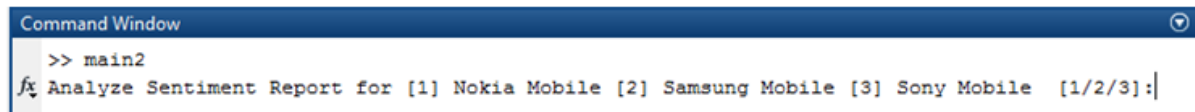
Sr. No.	System	Recall Value	Precision
1.	Modified LexRank	45.56%	-
2.	Proposed model	99.84%	98.18%

The proposed model and existing model results have been measured using the ROUGE system, which has described the proposed model's effectiveness and accuracy to calculate the results of sentiment analysis on the dataset collected from the social media platforms such as Facebook, Twitter, etc. The proposed model has also been evaluated for the precision value which is the probability of the positive predictive value (PPV). The proposed model has been evaluated as the way efficient method than the existing model of sentiment analysis and automatic product review summarization. The new text summarization reflects the real performance of the proposed model.

The result snapshots have been shown in the following section:

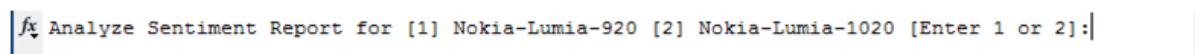
Run the Main file from matlab window

Command prompt shows:



```
Command Window
>> main2
fx Analyze Sentiment Report for [1] Nokia Mobile [2] Samsung Mobile [3] Sony Mobile [1/2/3]:|
```

Now pressing 1, 2 or 3, it's your wish for which particular mobile phone you want to see the reviews on. Let's say, press 1, then command window appears:



```
fx Analyze Sentiment Report for [1] Nokia-Lumia-920 [2] Nokia-Lumia-1020 [Enter 1 or 2]:|
```

Then, press 1 for Nokia Lumia 920 or press 2 for Nokia Lumia 1020. Let us say press 1

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

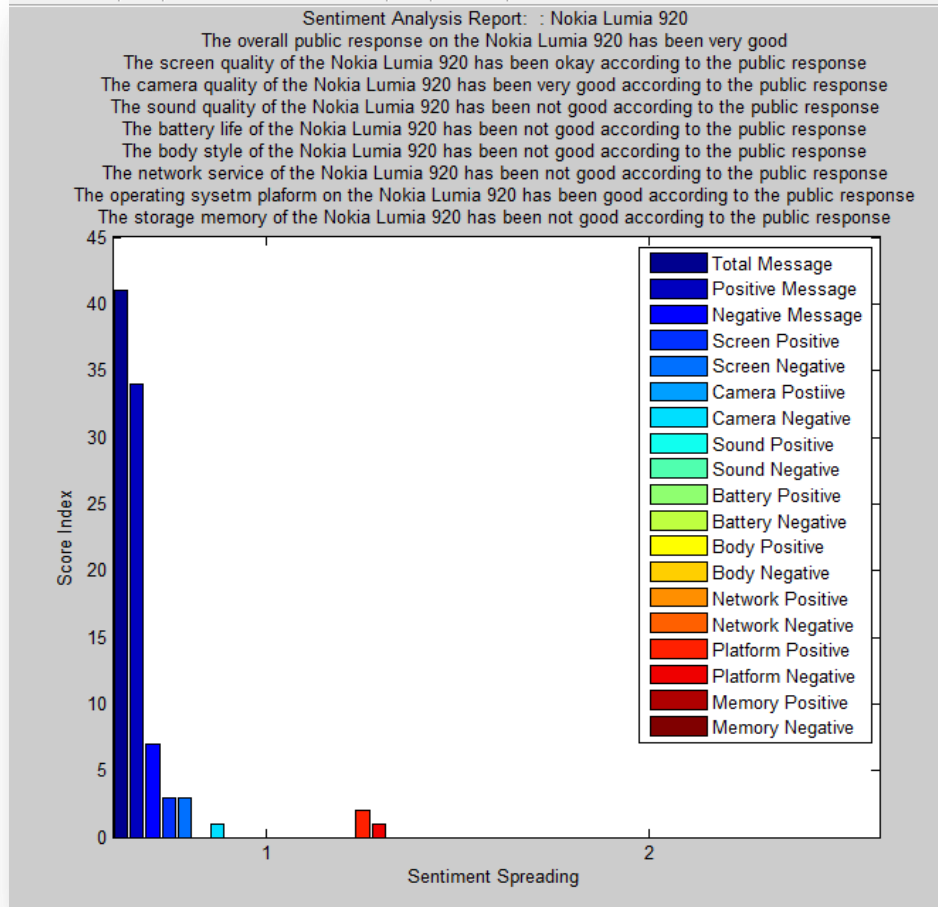


Figure10: Sentiment Analysis Report of Nokia 920

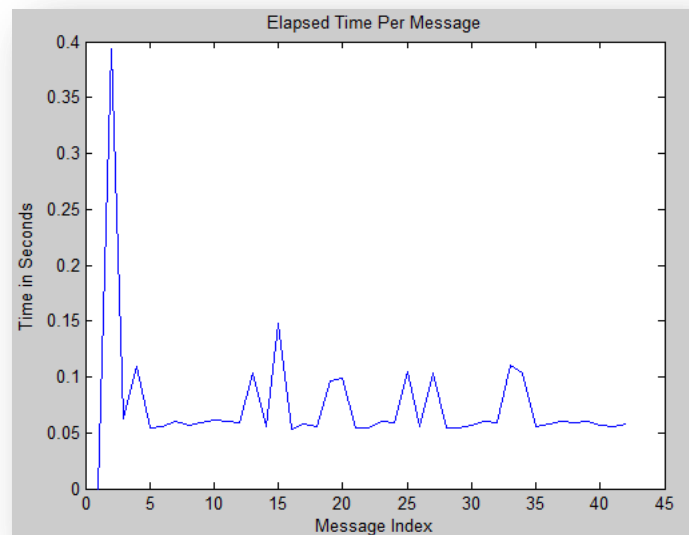


Figure11: Elapsed time per message of Nokia 920

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

Now, again run the main file to choose another model of Nokia i.e. for Nokia Lumia 1020. Press 2:

```
Command Window
>> main2
Analyze Sentiment Report for [1] Nokia Mobile [2] Samsung Mobile [3] Sony Mobile [1/2/3]:1
Analyze Sentiment Report for [1] Nokia-Lumia-920 [2] Nokia-Lumia-1020 [Enter 1 or 2]:2
```

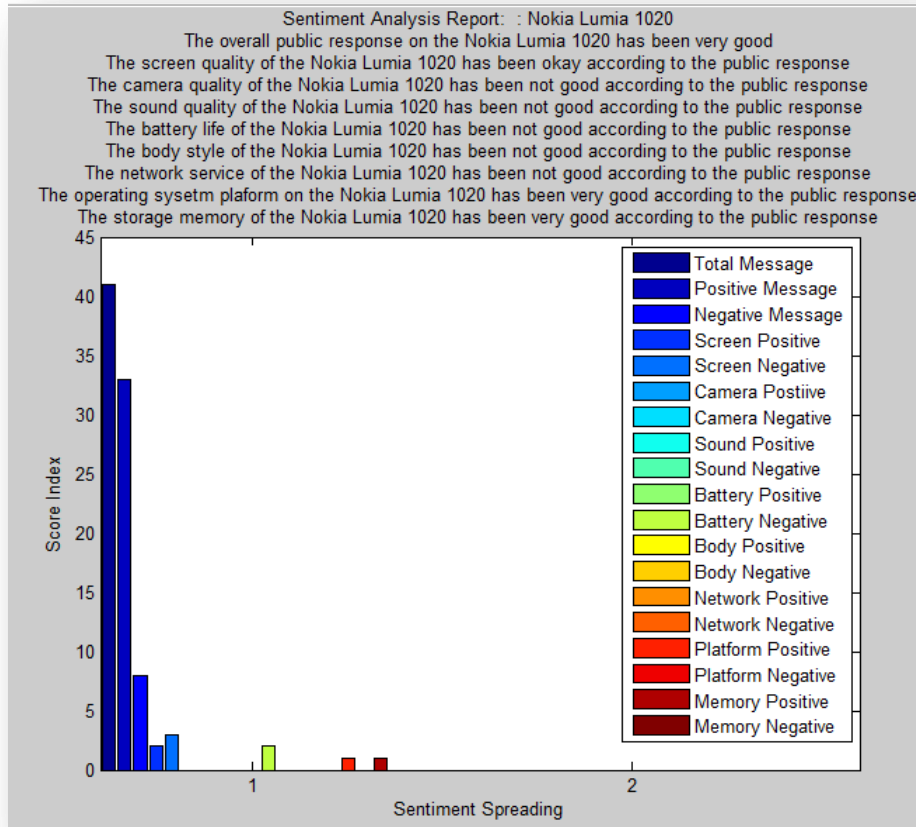


Figure12: Sentiment Analysis Report of Nokia1020

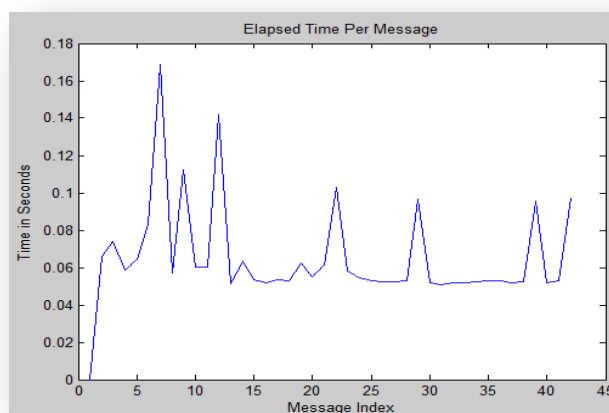


Figure13: Elapsed time per message of Nokia 1020

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

Now, analyze the report for Samsung phone. For Samsung press 2.

Press 1 for Samsung S6 Edge

```
>> main2
Analyze Sentiment Report for [1] Nokia Mobile [2] Samsung Mobile [3] Sony Mobile [1/2/3]:2
Analyze Sentiment Report for [1] Samsung S6 Edge [2] Samsung A5 [Enter 1 or 2]:1
```

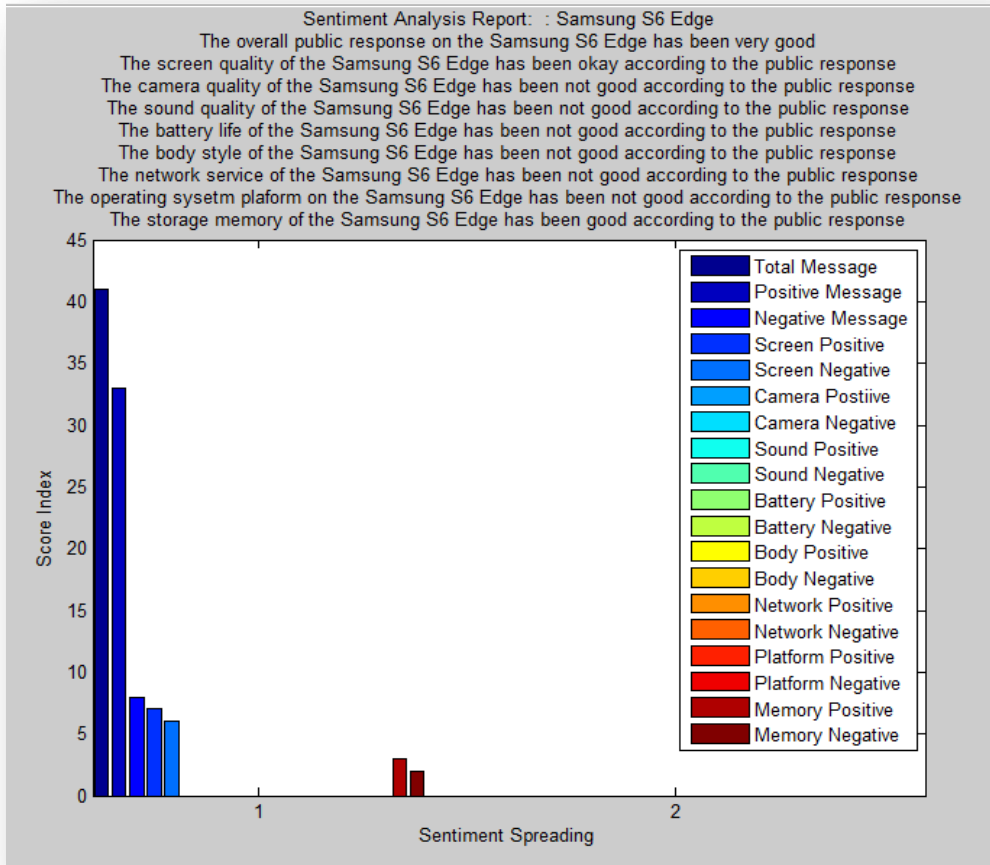


Figure14: Sentiment Analysis Report of Samsung S6 Edge

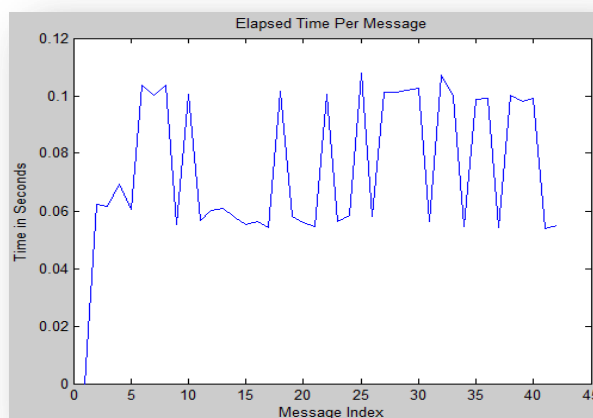


Figure15: Elapsed time per message of Samsung S6 Edge

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

Press 2 for Samsung A5

```
>> main2
Analyze Sentiment Report for [1] Nokia Mobile [2] Samsung Mobile [3] Sony Mobile [1/2/3]:2
Analyze Sentiment Report for [1] Samsung S6 Edge [2] Samsung A5 [Enter 1 or 2]:2
fx |
```

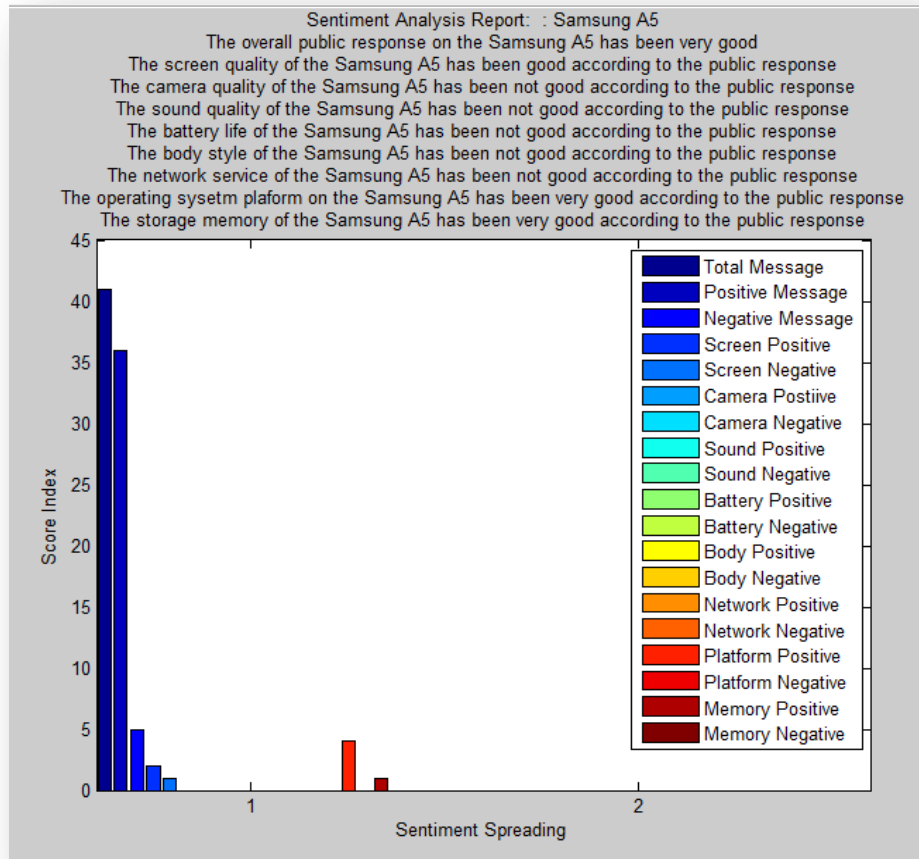


Figure16: Sentiment Analysis Report of Samsung A5

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

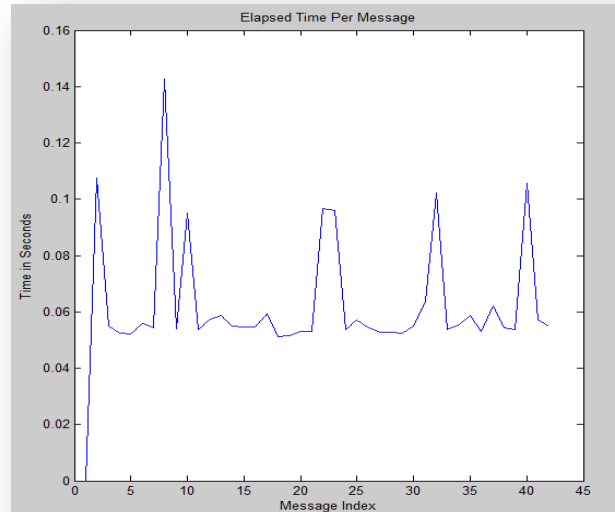


Figure17: Elapsed time per message of Samsung A5

Press 3 for analyzing Sony mobile.

Press 1 for Sony Xperia Z1.

```
>> main2
Analyze Sentiment Report for [1] Nokia Mobile [2] Samsung Mobile [3] Sony Mobile [1/2/3]:3
fx Analyze Sentiment Report for [1] Sony Xperia Z1 [2] Sony Xperia M4 [Enter 1 or 2]:1
```

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

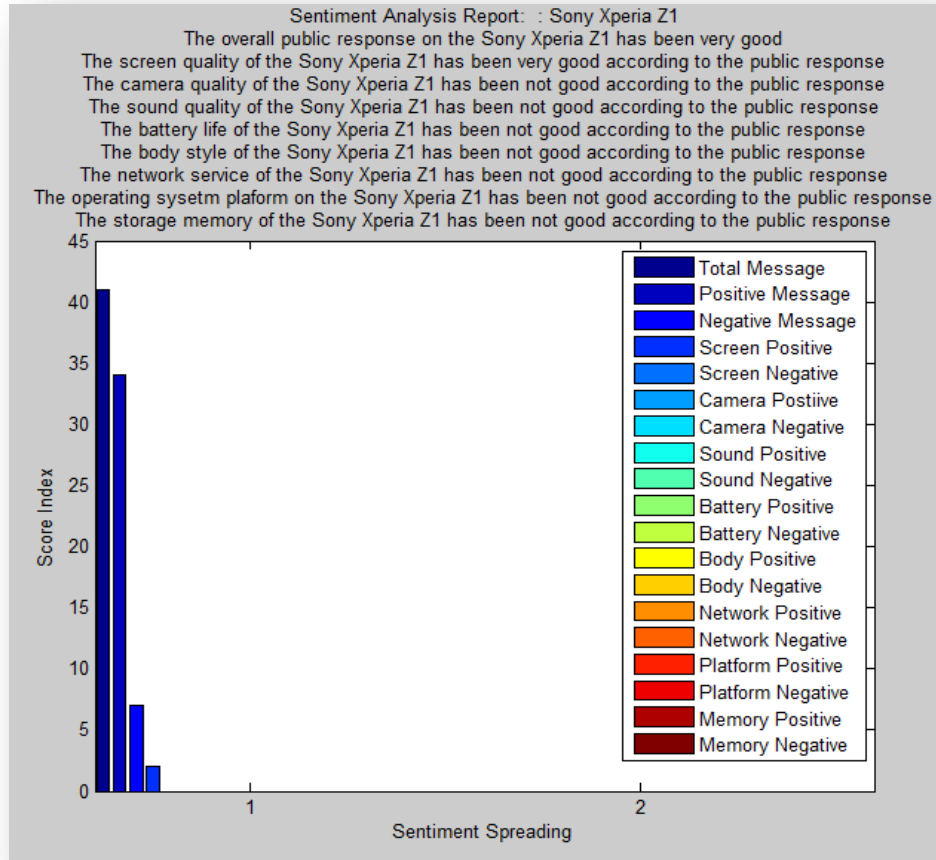


Figure18: Sentiment Analysis Report of Sony Xperia Z1

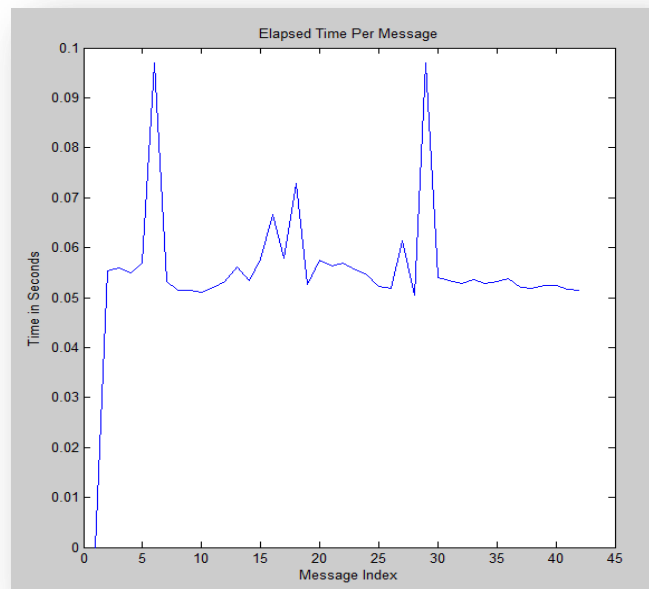


Figure19: Elapsed time per message of Sony Xperia Z1

Press 2 for Sony Xperia M4

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

```
>> main2
Analyze Sentiment Report for [1] Nokia Mobile [2] Samsung Mobile [3] Sony Mobile [1/2/3]:3
fx Analyze Sentiment Report for [1] Sony Xperia Z1 [2] Sony Xperia M4 [Enter 1 or 2]:2
```

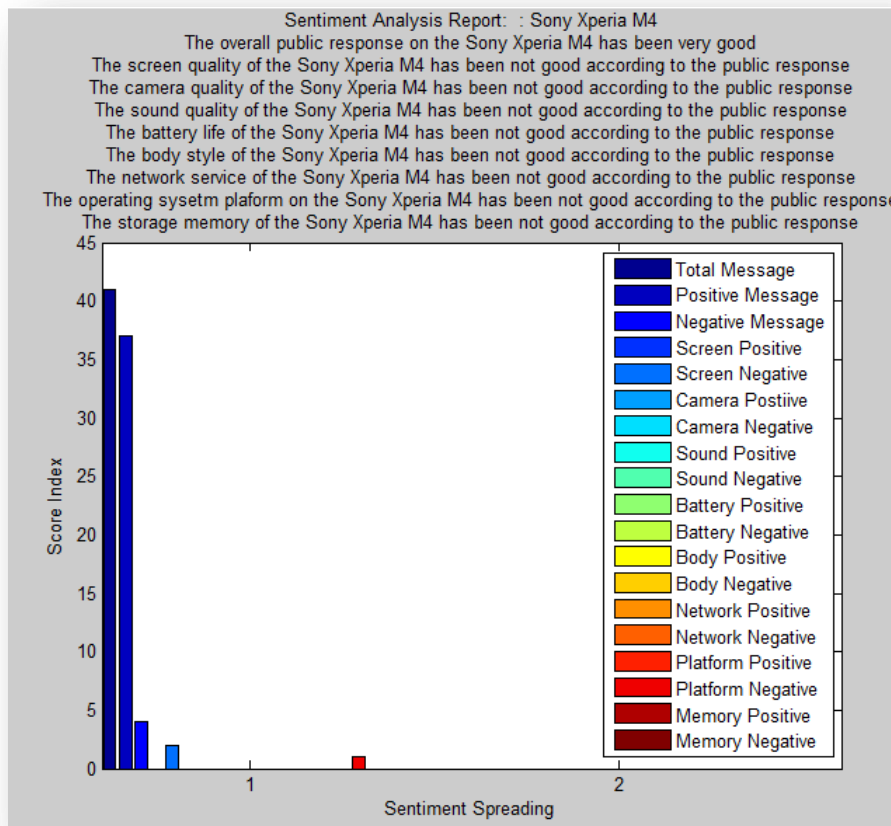


Figure20: Sentiment Analysis Report of Sony Xperia M4

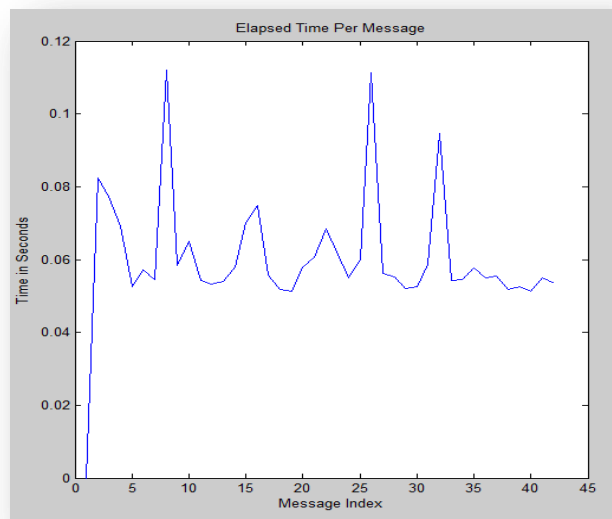


Figure21: Elapsed time per message of Sony Xperia M4

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

Compare Module: This module is all about comparison of features among the different models of different cell phone companies.

Say, a user wants to compare Nokia Lumia 920 with Samsung S6 Edge

```
Command Window
>> compare
Compare the Sentiment Report for the following
[1] Nokia Lumia 920 [2] Nokia Lumia 1020
[3] Samsung S6 Edge [4] Samsung A5
[5] Sony Xperia Z1 [6] Sony Xperia M4
Enter the first Choice:1
fx Enter the second Choice:3
```

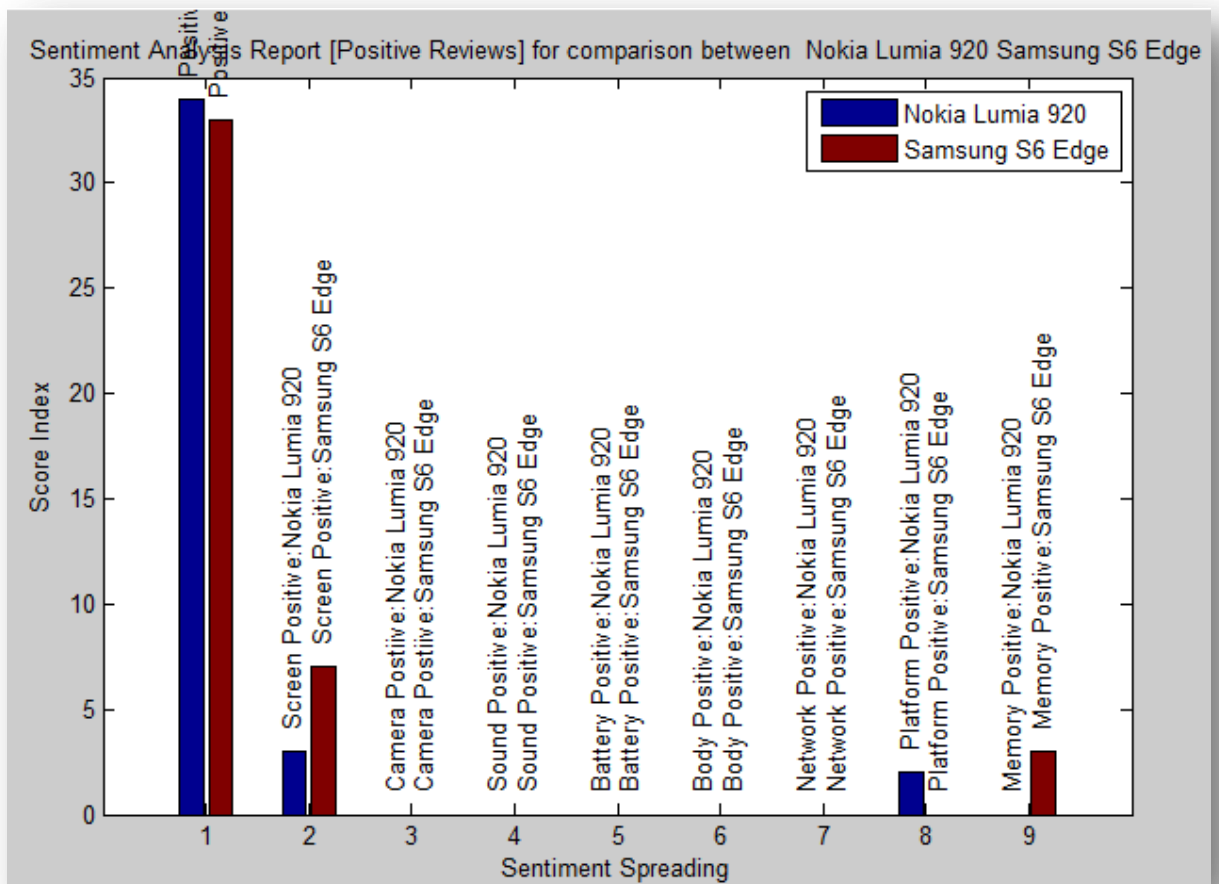


Figure22: Comparison between Nokia Lumia 920 and Samsung S6 Edge

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

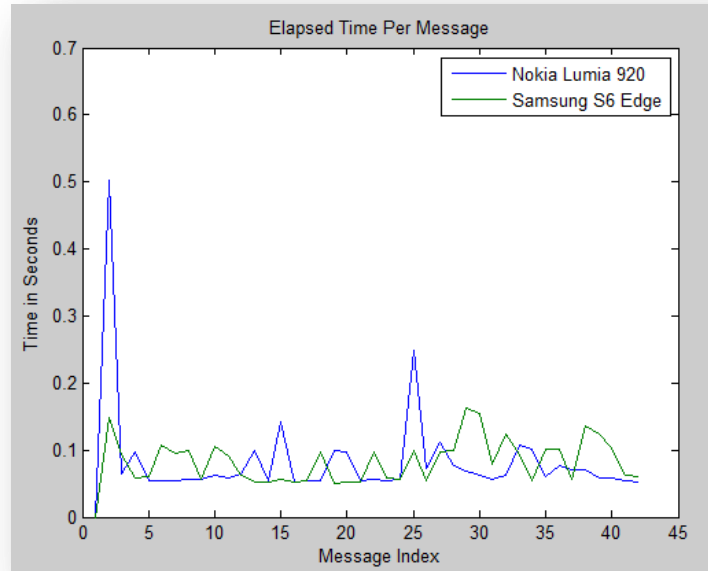


Figure23: Elapsed time per message

Now, let us say to compare between Samsung A5 and Sony Xperia Z1

```

Command Window
Compare the Sentiment Report for the following
[1] Nokia Lumia 920 [2] Nokia Lumia 1020
[3] Samsung S6 Edge [4] Samsung A5
[5] Sony Xperia Z1 [6] Sony Xperia M4
Enter the first Choice:4
fx Enter the second Choice:5
    
```

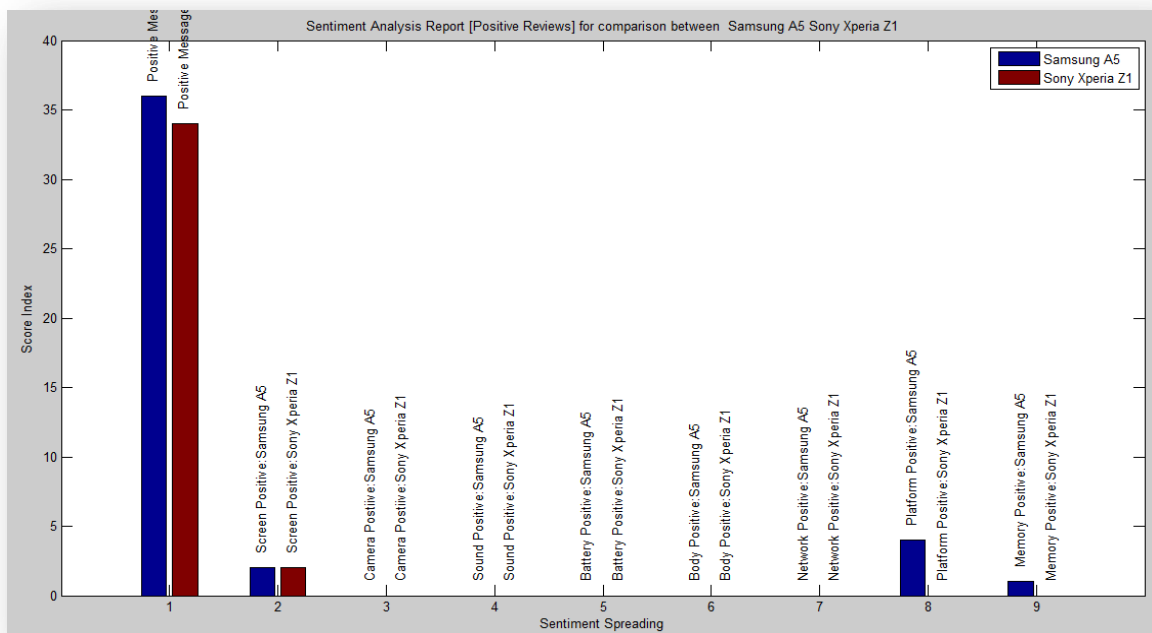


Figure24: Comparison between Samsung A5 and Sony Xperia Z1

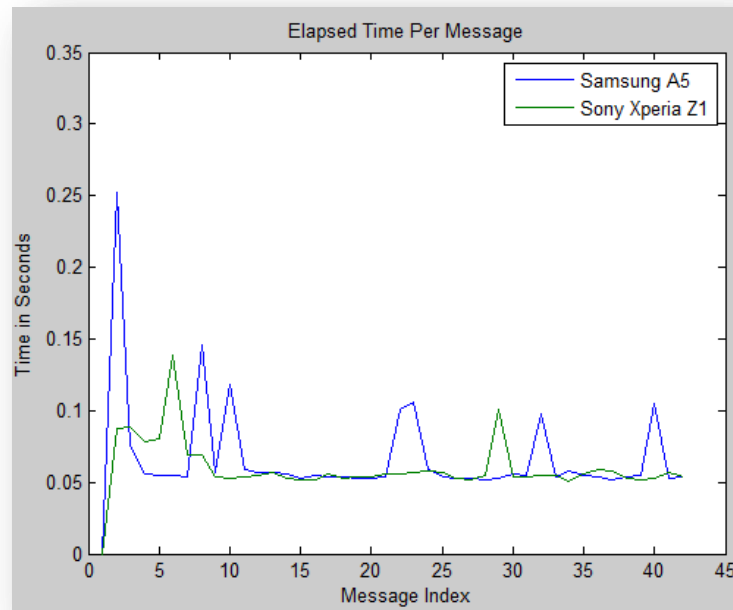


Figure25: Elapsed time per message

Module3: Comparison report on the basis of a particular feature asked by user.

Firstly, select the two models on which a user want comparison.

```

Command Window

Compare the Sentiment Report for the following
[1] Nokia Lumia 920 [2] Nokia Lumia 1020
[3] Samsung S6 Edge [4] Samsung A5
[5] Sony Xperia Z1 [6] Sony Xperia M4
Enter the first Choice:1
Enter the second Choice:4
    
```

After selecting two models for comparison, select the feature on basis of which a user want to select the one model among two selected models.

```

Compare on the basis of following features:
[1] Display [2] Camera [3] Sound [4] Battery
[5] Body [6] Network [7] Platform [8] Memory
Enter Your Choice: 1
    
```


SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

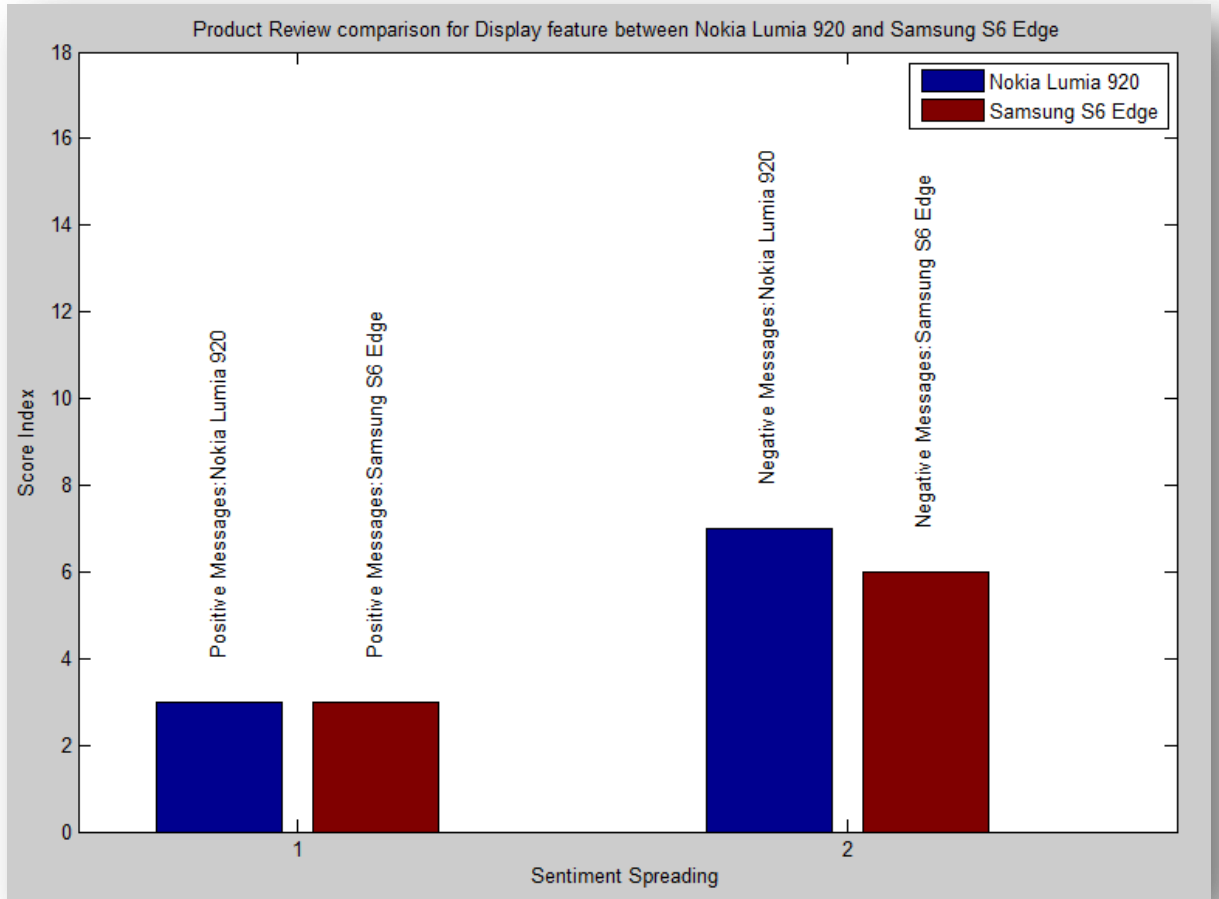


Figure26: Comparison between Nokia Lumia 920 and Samsung S6 Edge on basis of Display feature

CHAPTER 5

CONCLUSIONS AND FUTURE SCOPE

The product evaluation & summarization process consists of two major components: sentiment analysis & summarizer. The sentiment analysis component gives the sentiment spread in order to evaluate the user opinion on the product being evaluated using the input data. The proposed model has been designed in the different components for sentiment analysis and product review summarization. The proposed model performance has been measured in terms of Precision and Recall. Both precision and recall have produced the satisfactory results in terms of product review auto classification and automatic text summarization. The precision has been recorded near 98%, whereas the recall values have been measured at almost 99%. The existing system has been measured at almost 46% recall value, which is way lower than the recall produced by the proposed model. The proposed model has been also evaluated for its performance on the sentiment analysis. The sentiment analysis is the core system in the proposed model. Product review evaluation and automatic product review summarization depends upon the sentiment analysis report. The sentiment analysis report generates the emotion weights of the messages or reviews given by the users of the product specific. The specific product review classification is automatically done on the basis of product review ontology. The sentiment analysis system has been well tested for its performance on the various numbers of the product reviews. The system has been tested with 100, 1000, 10000 and other number options of the product reviews. The system has been proved its accuracy at nearly 96% for the sentiment analysis, which makes it the robust system.

In the future, the proposed model will be designed for multiple product review and product feature classification. Also the proposed model can be attached with some of the web source offering the product review API to classify the large amounts of data automatically. The proposed model can be enhanced for different N-gram and product feature classification.

CHAPTER 6

REFERENCES

Research Papers

- Ali Harb, Gerard Dray, Michel Plantié. (n.d.). Web opinion mining: How to extract opinions from blogs?
- Alok Ranjan Pal, Diganta Saha. (n.d.). An Approach to Automatic Text Summarization using WordNet.
- Amandeep Kaur, Vishal Gupta. (2013). A Survey on Sentiment Analysis and Opinion Minin Techniques. *Journal of Emerging Technologies in Web Intelligence*.
- Ana-Maria Popescu and Oren Etzioni. (n.d.). Extracting Product Features and Opinions from Reviews.
- Arti Buche, Dr. M. B. Chandak, Akshay Zadgaonkar. (2013). OPINION MINING AND ANALYSIS: A SURVEY. *IJNLIC*.
- Bakhtawar Seerat, Farouque Azam. (2012). Opinion Mining: Issues and Challenges (A survey). *International Journal of Computer Applications*.
- C. Lakshmi Devasenal and M. Hemalatha. (2012). Automatic Text Categorization and Summarization using Rule Reduction. *IEEE-ICAESM*.
- Dongjoo Lee, Ok-Ran Jeong, Sang-goo Lee. (n.d.). Opinion Mining of Customer Feedback Data on the Web.
- Dr. Ritu Sindhu, Ravendra Ratan Singh Jandail, Rakesh Ranjan Kumar . (2014). A Novel Approach for Sentiment Analysis and Opinion Mining. *ijetae*.
- E.PadmaLahari, D. K. (2014). Automatic Text Summarization with Statistical and Linguistic Features using Successive Thresholds. *2014 IEEE- (ICACCCT)*.
- Freimut Bodendorf, Carolin Kaiser. (n.d.). Detecting Opinion Leaders and Trends in Online Social Networks.
- G.Angulakshmi, Dr.R.ManickaChezian. (2014). An Analysis on Opinion Mining: Techniques and Tools. *International Journal of Advanced Research in Computer and Communication Engineering*.

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

- Hovy, Soo-Min Kim and Eduard. (n.d.). Automatic Identification of Pro and Con Reasons in Online Reviews. USC Information Sciences Institute, 1-8.
- Jiaming Zhan, Han Tong Loh, Ying Liu. (2009). Gather customer concerns from online product reviews – A text summarization approach. ScienceDirect, 1-9.
- Ladda Suanmali, Naomie Salim, Mohammed Salem Binwahlan. (2009). Fuzzy Logic Based Method for Improving Text Summarization. International Journal of Computer Science and Information Security, 1-6.
- Liu, Minqing Hu and Bing. (n.d.). Mining Opinion Features in Customer Reviews. Department of Computer Science University of Illinois at Chicago, 1-6.
- MAQBOOL AL-MAIMANI, NAOMIE SALIM, AHMED M. AL-NAAMANY. (2014). SEMANTIC AND FUZZY ASPECTS OF OPINION MINING. Journal of Theoretical and Applied Information Technology.
- Monika yadav, Pradeep Mittal. (2013). Web Mining: An Introduction. ijarcsee.
- MR. S. M. VOHRA, PROF. J. B. TERAIYA. (n.d.). COMPARATIVE STUDY OF SENTIMENT ANALYSIS TECHNIQUES. Journal of Information, Knowledge and Research in Computer Engineering.
- Periakaruppan Sudhakaran, Shanmugasundaram Hariharan and Joan Lu. (2013). Research Directions, Challenges and Issues in Opinion Mining. ijast.
- Pooja Kherwa, Arijit Sachdeva, Dhruv Mahajan, Nishitha Pande, Prashant Kumar Singh. (2014). An approach towards comprehensive sentimental data analysis and opinion mining. IEEE.
- Pravesh Kumar Singh, Mohd Shahid Husain. (n.d.). METHODOLOGICAL STUDY OF OPINION MINING AND SENTIMENT ANALYSIS TECHNIQUES.
- R.V.V Murali Krishna and Ch. Satyananda Reddy. (2012). A sentence scoring method for extractive text summarization based on Natural language queries. IJCSI.
- Raymond Kosala, Hendrik Blockeel. (n.d.). Web Mining Research: A Survey.
- SANJAY MADRIA, SOURAV S BHOWMICK. (n.d.). Research Issues in Web Data Mining. Center for Advanced Information Systems of Applied Science Nanyang Technological University. Singapore.
- Sebastiani, Andrea Esuli and Fabrizio. (n.d.). SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining.
- Seyed Hamid Ghorashi, Roliana Ibrahim, Shirin Noekhah and Niloufar Salehi Dastjerdi. (2012). A Frequent Pattern Mining Algorithm for Feature Extraction of Customer Reviews. IJCSI International Journal of Computer Science Issues, 1-7.

SUMMARIZING CUSTOMER REVIEWS – A TEXT SUMMARIZATION APPROACH

- Subhabrata Mukherjee, Pushpak Bhattacharyya. (n.d.). Feature Specific Sentiment Analysis for Product Reviews.
- Udo Hahn, I. M. (n.d.). The Challenges of Automatic Summarization.
- V. Bharanipriya¹ & V. Kamakshi Prasad. (2011). WEB CONTENT MINING TOOLS: A COMPARATIVE STUDY. International Journal of Information Technology and Knowledge Management.
- Walter Kasper, Mihaela Vela. (n.d.). Sentiment Analysis for Hotel Reviews. Proceedings of the Computational Linguistics-Applications Conference .
- Wesley T. Chuang, Jihoon Yang,. (n.d.). Extracting Sentence Segments for Text Summarization: A Machine Learning Approach.
- Y. Surendranadha Reddy, Dr. A.P. Siva Kumar, . (2012). An Efficient Approach for Web document summarization by Sentence Ranking. ijarcsee.
- YuanbinWu, Qi Zhang, Xuanjing Huang, LideWu,. (n.d.). Phrase Dependency Parsing for Opinion Mining.

Websites:

- <http://brnr.me/social-sentiment-sentiment-analysis/>. (n.d.). Retrieved from brnr.me.
- http://en.wikipedia.org/wiki/Sentiment_analysis. (n.d.). Retrieved from wikipedia.
- <http://smartdatacollective.com/raminuseir/191466/strange-uses-sentiment-analysis>. (n.d.). Retrieved from smartdatacollective.com.
- <http://www.bitext.com/bitext-api-2/sentiment-analysis-2.html>. (n.d.). Retrieved from bitext.com.
- <http://www.marketingdonut.co.uk/marketing/customer-care/understanding-your-customers/how-to-use-customer-feedback-to-improve-your-business>. (n.d.). Retrieved from <http://www.marketingdonut.co.uk>.
- <http://www.pcworld.com/article/189525/sss.html>. (n.d.). Retrieved from <http://www.pcworld.com>.
- <http://www.slideshare.net/jaganadhg/opinion-mining-and-sentiment-analysis-issues-and-challenges>. (n.d.). Retrieved from slideshare.net.

Books:

- Ani Nenkova, Kathleen McKeown. (n.d.). Automatic Summarization.

CHAPTER 7

APPENDIX

POS Tagging: POS Tagging stands for part-of-speech tagging also known as grammatical tagging. It is the process of marking up a particular word in a text i.e. corpus corresponding to a particular part of speech depending on its definition and its relationship with the adjacent and related words in a phrase, sentence, or paragraph.

Stop word removal: It is the process of removing stop words. Stop words are words which are used in formation of sentence but do not have much of importance. Stop words include a, an, the, to, for, etc.

Word stemming: Word stemming is the process used to remove prefixes and suffixes of each word. It means converting each word into its root. For an example, fisher, fishing words are stemmed as fish.

Opinion Mining or Sentiment Analysis: Sentiment analysis is a process to determine the outlook of the reviewer regarding some matter.