Sentiment Analysis and Mining of Opinions

Surbhi Bhatia, Manisha Sharma and Komal Kumar Bhatia

Abstract Now days the way of expressing opinions on certain products that people purchase and the services that they receive in the various industries has been transformed considerably because of World Wide Web. Social Networking sites fascinate people to post feedbacks and reviews online on blogs, Internet forums, review portals and much more. These opinions play a very important role for customers and product manufacturers as they tend to give better knowledge of buying and selling by setting positive and negative comments on products and other information which can improve their decision making policies. Mining of such opinions have focused the researchers to pay a keen intention in developing such a system which can not only collect useful and relevant reviews online in a ranked manner and also produce an effective summary of such reviews collected on different products according to their respective domains. However, there is little evidence that researchers have approached this issue in opinion mining with the intent of developing such a system. Our work will focus on what opinion mining is the existing works on opinion mining, the challenges in the existing techniques and the workflow of mining opinions. Consequently, the aim of this chapter is to discuss the overall novel architecture of developing an opinion system that will address the remaining challenges and provide an overview of how to mine opinions. Existing research in sentiment analysis tend to focus on finding out how to classify the opinions and produce a collaborative summary in their respective domains, despite an increase in the field of opinion mining and its research, many challenges remain in designing a more comprehensive way of building a system to mine opinions. This chapter addresses the problem of how to classify sentiments and develop the opinion system by combining theories of supervised learning.

S. Bhatia (✉) · M. Sharma
CS department, Banasthali University, Vanasthali, India
e-mail: surbhibhatia1988@yahoo.com

M. Sharma
e-mail: manishasharma8@gmail.com

K.K. Bhatia
CSE Department, YMCA University of Science and Technology, Faridabad, India
e-mail: komal_bhatia1@rediffmail.com

N. Dey et al. (eds.), *Internet of Things and Big Data Analytics Toward Next-Generation Intelligence*, Studies in Big Data 30,
DOI 10.1007/978-3-319-60435-0_20

Keywords Mining · World Wide Web · Opinion · Sentiment analysis · Supervised learning

1 Introduction

Opinion Mining is closely associated with the Mining area. We can consider opinions to be a feeling, review, sentiment or assessment of an object, product or entity [1]. Mining refers to as extracting some useful information and representing it in the form of knowledge in an understandable structure as patterns, graphs etc. [2]. The two important aspects of Opinion Mining are Information Extraction and Question Answer [3]. Opinion mining is a technique of extracting knowledge automatically by taking opinions from others on the specific topic or problem [4]. Sentiment analysis can be used as a synonym for opinion mining. The increase in the growth of Social media has made the people dependent on search engines and Information retrieval systems. This dependency has created the need to develop such a system that will help people in making the correct decision of purchasing products by analyzing the positivity and negativity score of the opinions respective to the particular domain.

Opinion mining has numerous applications in purchasing and marketing products and services, setting up policies and making decisions, marketing and trading etc. We will explore the trends in the mining industry; in particular, how online reviews are reflecting the major shift in the decision making policies of customers. The various researchers like [5–8] have discussed briefly the meaning of opinion mining, its techniques, tasks, sentiment analysis and other challenging aspects in summarization of opinions. Penalver-Martinez discussed opinion mining in semantic web technologies [9]. A novel technique of assessing product reviews is presented by [10]. The aspect based classification and summarization of reviews is depicted by proposing a novel model which is represented with a generic tool which does not require any seed word or pre requisite knowledge about the product or service [11]. Mfenyana has proposed a tool to collect and visualize data from face book [12]. Features based summarization is discussed in detail by Bafna [13]. Sentiment analysis done by using learning techniques is evaluated and discussed diagrammatically [13]. The aspect based opinion mining by combining hybrid theories are discussed and results are evaluated by effective measures using search strategies [14]. Pang worked out the effectiveness of sentiment analysis problem with the application of machine learning algorithms [15]. Hemalatha developed the sentiment analysis tool that focuses on analyzing tweets from social media [16]. McDonald identified sentiments at all the levels with its subcomponents [17]. Liu and Chen presented a detailed analysis of twitter messages by making use of three metrics [18]. Bhatia presented the design and implementation of opinion crawler [19]. We will therefore design a novel system that is dynamic in nature containing less complex algorithms and promises to generate more effective and structured opinions considering its aspects.

Most previous research in opinion mining and sentiment analysis focuses on a specific content. Thus, there is little emphasis on a coherent structure—no integrative framework to mine opinions is proposed. Moreover, there is little attention given to conceptualize the structure of mining opinions. It has also been observed that many institutes are consistently doing research in Opinion mining area which has been listed as one of the prime means of creating enormous amount of online reviews and feedbacks of different products. The following section focus on the systematic literature review with the basics of Opinion mining.

The goals of the chapter are as follows:

(1) To define and explain Opinion Mining.
(2) To find the gaps in the existing definitions and frameworks.
(3) To explicate the need of information retrieval in Opinion Mining.
(4) To propose the structure using Opinion Mining.

The chapter has been structured as follows. The definition, existing challenges, literature review and the background has been explained in Sect. 2. Section 3 examines the framework design and explains all the modules in determining opinions. The analysis of the proposed work and results are evaluated in Sect. 4. The summarization is discussed in Sect. 5 and the scope for the future is concluded.

2 Literature Review

The Opinion mining is the crucial part of the Data mining, a significant part of our paper addresses the subject. In the previous work done, the various researchers examined it from a narrow scope. An in-depth study is required to present an overview of what has been published on opinion mining. So a need to determine and identify the areas in which detailed research is required in order to meet the demand required by the present and future day systems being built is explored in our work. A systematic mapping involves going into the depth of the literature and examining the nature, content and quality of the published papers in this domain. The grouping of primary studies is done to produce a view of the specific research area by the systematic mapping of the related studies. The systematic literature review process is necessary to ensure that the further innovative research can be carried out by the various researchers, practitioners and also to aim at gaining an overall idea of the distinct researchers in this field.

The present study intends to summarize the existing evidence in the opinion mining area. The review is undertaken in accordance with the predefined search strategy as proposed by Kitchenham [20]. The established of such a systematic review helps in minimizing the researcher bias. Few research questions are listed which seem to be a part of discussion in the literature review section given below. Section 1.2.2 defines the paper literature extraction strategy containing resource archives.

2.1 Research Questions

The main target of this assessment is to report the research done in the field of opinion mining. The following questions are discussed which reflect our purpose as under:

RQ1: How is the term 'opinion mining defined'?
RQ2: What types of studies are published by the abundant researchers?
RQ3: How components are inter-related and what are the different techniques used in classifying sentiments?
RQ4: How strong is the association between opinion mining and social media?

The relevant literature is a twofold process, i.e. searching through a keyword in the scientific libraries and second is the collecting papers from the workshop series. It consists of the following steps.

- Gathering information about the literature: This is the combination of the digital library and distinct sources such as ACM, Springer, and IEEE, Science Direct etc. which includes the search based on titles, keywords, and abstracts. The peer review process is considered for study selection in order to attempt for a broader prospective, so the restriction on the primary studies must be avoided.
- Applying the selection/rejection method: The paper that explores the relevant study regarding Opinion mining is taken into account to be selected. The papers that are not related i.e. other than mining opinions are rejected.
- Verify the rejected papers: The combination of the different keywords must be avoided. Only the terms "mining" and "opinion(s)" in the field's title, abstract, keywords are evaluated in the literature.
- Verify included papers: This is manually done by reading the abstracts and conclusions. The selected literature resulted to produce outputs that are directly or indirectly related to mining opinions.

2.2 Analysis

Based on the above mentioned research questions and the analogies extracted for RQ1, RQ2 and RQ3 with respect to RQ4, thorough analysis has been accomplished to achieve the goal.

The thorough literature survey has been done in order to present the in-depth study and the results are presented in the following section according to the questions listed above.

2.2.1 Defining Opinion Mining

Human feeling and emotions can be described and entities, events and properties are evaluated using a subjective approach that can be taken as an Opinion. Opinion mining can be considered as a process of collecting reviews that are posted online on various Social Networking sites. This definition, which is most prevalent, has been used by a number of researchers. This definition considers mining as a bigger term which contains many sub definitions.

According to another definition, to identify and abstract subjective facts and material from text documents can be accomplished by applying Opinion mining technique. There are various terms which can be used exchange ably with Opinion Mining like Review mining, Sentiment analysis, Appraisal Extraction as suggested in paper [21]. Considering the hierarchical structure, Web content mining list Opinion mining as part of it which basically comes under the category of Web Mining [22]. Data mining and World Wide Web lies in close proximity with one another. Several research communities consider web mining research an important resource of research since it lies at its junction. Data mining and World Wide Web lies in close proximity with one another. Several research communities consider web mining research an important resource of research since it lies at its junction. Each term is explained in detail as follows:

Data Mining (DM): The other name of DM is Knowledge Discovery. It is an important process which deals with analysis of data from various perspectives and summarizes it into valuable information that is useful for companies and individuals.

Text Mining (TM): It is analysis of data present in natural language text. It aims at getting fine information from text through means such as statistical pattern learning.

Web Mining (WM): It can be stated as discovering and analyzing useful information from World Wide Web. It is basically categorized into three types as Web usage mining, Web content mining and Web structure mining.

Web Usage mining: It is the process of discovering what users want to see on the Internet. The data can be of the textual form or in multimedia form by making use of user logs.

Web content mining: It targets the knowledge discovery including multimedia documents like images, video and audio. It scans the contents of web page deeply and extracts useful information.

Web Structure mining: It focuses on analysis of the link structure of the web and generate structural summary about the Web site and Web page.

The evidence of the above is database, information retrieval, and Artificial Intelligence [23]. The complete hierarchical model of data mining is diagrammatically explained in Fig. 1.

We can define Opinion Mining as a set of text documents (T), entity which has opinions, further determining attributes of the entity in text $t \epsilon T$ and comments to be distinguished in the form of positive, negative or neutral classes [24]. It is one the most appropriate definitions as it considers different components evolved and

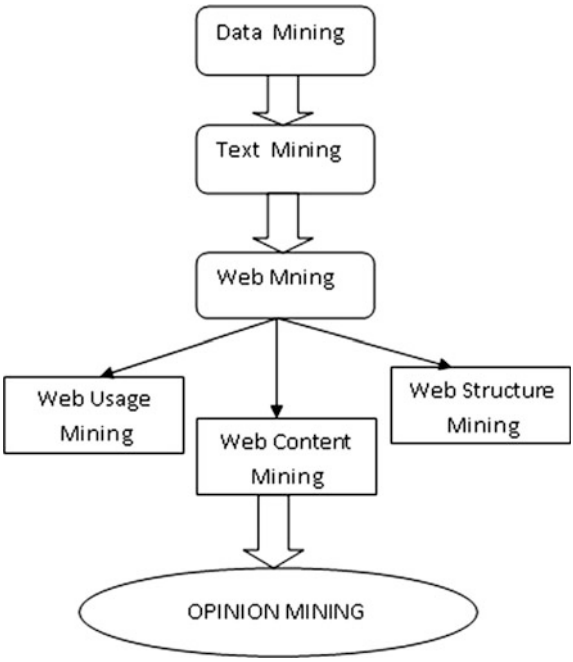

Fig. 1 Simplified model of data mining

their interactions with each other. The resonance of this definition with Information Retrieval is hard to ignore. In fact, the most important attribute of the definition of Opinion mining should be its ability to encompass varied components developed for the same product, working together for a common task. The definition should be open to ascertain the effect of components on the target audience and vice versa.

According to Bing Liu [6], the area that helps in analyzing person's opinion, emotions, attitudes, feelings towards various entities including products, services etc. is referred to as Opinion mining.

As discussed in his earlier works, Opinion Mining has been categorized into three levels;

- Document Level: Distinction is done taking into consideration the whole opinion document into positive category or negative category [15, 25].
- Sentence Level: This level goes into a deeper level than document, i.e. it distinct the complete sentence into positive, negative or neutral category. No opinion comes under neutral category. It is also related to subjectivity classification [26].
- Feature Level: A more detailed classification is aspect level performs advanced and more quality based analysis. Feature level is a synonym of Aspect level [27]. The features or views are taken into account at a deeper level than considering documents, paragraphs, sentences, clauses or phrases which come under language constructs. Opinions will be based on the target or object which will be either positive or negative.

According to our definition, Opinion Mining is a collaboration of reviews working as a prime component and interaction with the Public media for advertising and selling their goods and services. Maintenance of the relationship with the components is required which is underlined by the common platform by interchanging information, resources or artifacts.

It may be noted that this definition takes into consideration almost all the factors mentioned in the previous definitions. However, with respect to the E-Commerce, the author sees Opinion mining as a set of businesses which is somewhat true. Businesses run and interact with the social media with a sole purpose of making profit. If all the components of a particular Opinion Mining compete for individual profits, then the definition will be correct. However, in our view this is not the case. The components of a Opinion mining perform the task they are required to perform. They might not be intending to make the profit but just contributing towards the goal of the overall mining system. In this sense, the Information Retrieval analogy would be more appropriate than the businesses one.

2.2.2 Studies by Various Researchers

This section gives a brief of works given by various researchers. The study of Opinion mining is gaining importance, from its inception in 2012. It is evident from the increase in the number of papers published. It, therefore, becomes important to understand what exactly Opinion mining is. For any evolving subject a divergence of views is necessary and is healthy for evolution. It is not essential to follow the previous definitions although they might show the way. The way might have much segregation and there is no reason why perspective segregation should be left.

The paper [28] discusses the properties of social networks associated with opinion mining. It explores a new opinion classification method which a variant of semantic orientation and presents the results by testing the algorithm for accuracy on data sets from real world. The future work concentrates on the response of opinions in text and improving the performance of the algorithm by creating an active learning strategy.

The paper [29] presents an outline of opinion mining and sentiment analysis clearly with the data sources, components and tools. The levels of sentiment analysis with its techniques are discussed with examples. The paper [30] discusses how link based graphs can be applied on the newsgroups which helps in classifying objects in different categories and hence are considered more accurate. The future work concentrates on text information by inculcating linguistic analysis for improving the accuracy. The paper [31] discusses the importance and functionalities of different types of mining areas in social networks. It emphasizes on how opinion mining and sentiment analysis can be studied to gather knowledge which can promote business, political scenarios and other areas. It promises to deploy the techniques which can be applied in real world applications. The paper [32] presents a model based on opinion based graph which focuses towards content oriented domain. It allows observing the evolution of the opinion inside a discussion.

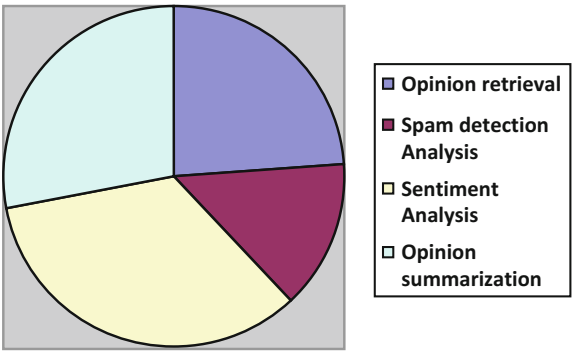

Fig. 2 Distributions of primary studies by research type

The comparison between the existing user based graph approach and the proposed opinion based graph is illustrated. The future work gives a summarized picture of combining user based graph and opinion based graph for analysis. It also explores how opinion changes over time. The paper [33] focuses on achievement of the opinion mining levels. The problems faced in sentiment analysis for reviewing a product are discussed in order to provide a summarized framework of opinions. Set of reviews and set of known expressions related to features which are taken from the reviews are provided as an input for Opinion Mining. The paper [5] presents a systematic literature review of opinion mining techniques and methodologies and also explores the existing gaps in the field. The author promises to improve the performance of the model and work on issues in sentiment analysis. The major limitations that need to be considered are language issues, dealing with negative terms and finally summarize the opinions on the basis of aspects etc. The paper [34] proposes a new framework for opinion mining with respect to content analysis in social media. It discusses the three prime components to track opinion data online. The further research focuses on the policy making issues through social media sources. Precise learning algorithms related to quantification perform sentiment analysis by conducting experiments on diverse tweets datasets. A naive way to tackle quantification is done by classifying each unlabelled item independently and computing the fraction of such items in the class is explored in the paper [35].

After going through the keen analysis of the various papers on Opinion mining, we classify the papers according to the study type. The major research has been done under the category of Sentiment analysis, followed by summarization, followed by Opinion retrieval and last is spam detection. Figure 2 shows the frequency of major studies by research type. Chiefly the distributions are listed.

2.2.3 Components, Techniques, Applications of Opinion Mining

Automatically identifying the opinions from documents and presenting it to the user is the prime core task of opinion mining [36, 37]. The task has been defined in terms of extracting public's opinion based on the features of products. These

components are distinguished in the below listed three categories under Opinion Mining:

- Opinion Holder
- Object (Entity/Target)
- View.

Considering the example:

I purchased LG Television three days ago. I loved watching it. The picture quality was great. It contained smart television feature also. The wi-fi slot enables me to watch my choice movies. However my brother was angry as it did not have the 3-D feature. He also thought that the TV was expensive, so I should return it.

There are several opinions in this review. Talking about the view: Sentences 2, 3, 4 and 5 convey optimistic response and sentences 6 and 7 convey pessimistic response. There are various targets in the above example. The target is phone in sentence 1, picture quality in sentence 3, wifi slot in sentence 5 and price in the last sentence. The holder is author in sentences 1, 2, 3, 4 and 5. But in the last two sentences i.e. 6 and 7, the holder is brother.

Policy making which is a joint decision can be easily obtained by making use of Opinion mining applications. It is useful in making sense of thousands of interventions [38]. Search engines, Question answering systems and other practical applications are making extensive use of Opinion mining. The interaction between humans and machines has also been improved with it [3]. The primary areas where Opinion mining can set a major impact are:

- Purchasing and Quality improvements in products and services by placing ads in user generated content.
- Decision making and Policy setters by tracking political topics.
- Automated content analysis which provide general search for opinions.

In general, Opinions can be considered as an image of certainty as it is dependent on recognizing difficulties by paying attention through listening rather than by questioning. To classify opinions into categories like positive, negative and neutral, we have a large no. of sentiment detection approaches. Sentiment detection is closely associated with Natural Language Processing and Information Retrieval. Focusing on the uses of machine learning which mainly belong to supervised classification and text classification techniques in case of Sentiment Analysis. So it is called as supervised learning. On the other hand, unsupervised classification belongs to semantic orientation approach. Both of these two approaches necessitate working on a data set which involves training with pre worked examples. Thus, classification requires training the model [5].

Training data and Test data are involved in a machine learning based classification. Distinguishing features are required to be understood by the automatic classifier while working on training data set in text documents. The performance is validated using test data set of the automatic classifier.

Text categorization mainly involves supervised learning methods such as Naive Bayes, Maximum entropy, Support Vector Machines, Winnow classifier, N gram model etc. [39–43]. Unsupervised learning makes use of available lexical resources. In this approach, input given is a set of unknown labels (not known during training) and output comes as classified data. Clustering is a good example of the above technique. Word lexicons which appear in sentences containing opinions are compared and sentiment words are determined for classifying. Part of speech tagger, software which works by allocating parts of speech to every word that appear as an input in the text is widely used in determining sentiments. Corpus based or dictionary based techniques are used in this type of technique.

3 Proposed Work

An integrative framework is discussed in this section that draws on previous research, extracts, classifies and summarizes the opinions posted online on various social networking sites. We argue that such a framework is an organizing and unifying structure that accounts for mining opinions in a more general and easier way. Such an integrative framework is temporal independent and periodic in nature as the crawler that we will design to retrieve opinions regularly updates itself. Moreover, the work we will present in building a crawler, there exists no single theoretical perspective that can support such a platform that can extract relevant and quality reviews ignoring the spam reviews but also periodically update the sites by calculating the timestamp of the URL's. The work will also include to classify the sentiments by using different supervised learning techniques and finally to collate all the opinions to produce the one liner summary. The final ranking of the opinions is also done using different algorithms. As initial analysis was unable to find evidence suggesting that their exists such a unique system of summarizing opinions, we hypothesize that a set of different learning techniques will account for classifying opinions which will form the groundwork for explicitly consolidating them and to present the final opinion.

Our emphasis will be on describing the opinion system as a framework that can be used to present ranked opinions according to different domains and describe the flow of mining opinions. We begin by outlining the meaning of opinion mining, its advantages, information retrieval system, and sentiment analysis with a focus on the evaluation by using various search strategies. Then, we present a brief overview of the system, and discuss its modules and components one by one. The modules describe the overall flow of mining opinions under various domains that can be measured and evaluated in information retrieval search strategies. Hence, the combination of modules will be used to construct the framework and will provide the final summarization of ranked wise opinions, which in turn is expected to give the final results in the ranked manner in the form of opinions. More importantly, this system can be used to define the development of a complete architecture to mine opinions, providing a description of each component, including its definition

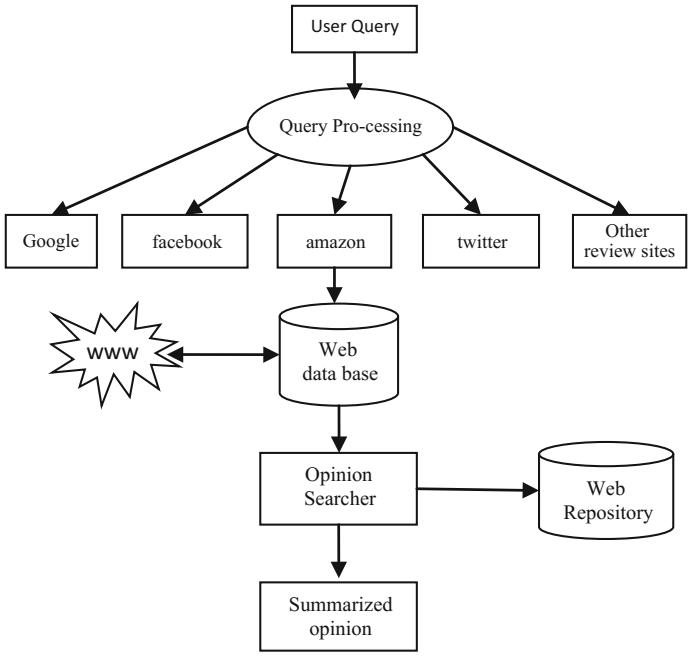

Fig. 3 Framework design

and behavior. Without a detailed description of every module, it would be difficult to design how the components would look and behave in mining opinions. On the other hand, with a set of definitions and behavior descriptions, more informed decisions can easily be made by people in whether to buy a certain product or not. The framework design is shown in Fig. 3.

The explanation of every module is given as follows:

- User Interface: It is basically an interface that enables users to write their queries for getting the output as opinions.
- Opinion Retriever: Opinion retriever contacts the World Wide Web to look into the Social Media, E-forums and other sites to collect reviews and feedbacks. The HTML source is parsed by checking the keywords such as comment, review summary, opinion in the div class or span class. Our earlier works develop such a crawler [19] that periodically updates reviews and present the most recent opinions with the date and time which fulfills our purpose of having relevant reviews which is time and temporal independent. The architecture for the Opinion Retriever is given in Fig. 4.
- Opinion Scheduler: This module will pick the URL from the queue of the URL database. If some URL's are specified as seed URL's, then it will prioritize and choose that URL from amongst all, otherwise it will follow the LIFO policy of

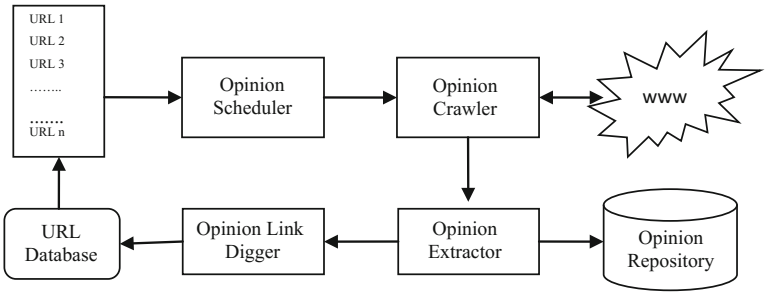

Fig. 4 Opinion retriever

the queue. It will first check the avail list of the URL buffer. If not empty, will traverse the buffer.

- Opinion Crawler: This module will fetch the URL from the buffer and will check whether the URL is of the review site or not. The code is taken and a structure is developed in memory which computer understands. HTML code is taken and relevant information like heading, paragraphs, links and hyperlinks in the page are taken to search for the opinions from the world Wide Web.
- Opinion Extractor: It will search for the specific tags from the HTML or XML pages to extract only the review content from the parsed HTML code by using basic string processing and index the opinions and store them in the repository.
- Opinion Repository: It is a storage area for the final relevant opinions found from the Opinion Extractor.
- Opinion Link Digger: This module will search the links and extract them from that particular web page in the depth first search manner and will add them in the URL queue associated with the URL Database.

The algorithm for the above is given in Table 1.

- Spam Detector: Spam can be referred to as an area of accomplishing search in text documents with the sole purpose to find irrelevant and fake opinions which can be regarded as useless. These reviews are created for the sole purpose of advertisement or deceiving the customers falsely. Machine learning approach has been widely studied in spam detection which has already been compared and studied [44]. This module will look for the spam reviews and filter out the relevant ones from the irrelevant ones. Since, words are repeated many a times in the fake reviews. Keyword stuffing is a technique which is used in our work by checking the review sentences with the keywords used by the spammers. Also, the threshold is set which checks to find if the length of review is more than expected.
- Tokenizer: This module will pick the relevant opinions one by one and will tokenize the opinions into individual sentences after removing the stop words and transforming the tokens from plurals to singulars. This will be done by using the automatic POS Tagger.

Table 1 Opinion retriever

char *text[], ch;
*text[] = {“list of URLs};
If(*text[i] ==NULL)
{
Printf(“no URL left”);
Exit(1);
}
While(ch ==gets(*text[])! = EOF)
{
If(found seedURL)
{
Prioritize URL;
}
else
Pick URL from *text[];
}
Fetch page;
Parse Page;
Resolve DNS;
Make threads for HTTP Socket;
If(!visible page content)
{
Break;
}
If(robots.txt exists)
{
If(.disallow statement exists)
{
Continue;
}
}
extract the opinion
If(meta tag found)
{
If(div class!! Span class)
{
If(review content!!comments found)
{
Extract sentence
}
}
}

(continued)

Table 1 (continued)

Store in opinions repository
}
Extract URLs;
Normalize URls;
If(already visited URL)
{
Remove from *text[])
}
Else
Add URL to *text[];
}

Table 2 Opinion categorizer

do
{
while (wo_i ! = NULL)
{
for(x = 1;x < n;x ++)
{
if (wo_i == T_d)
return +1 and −1
else
overall polarity = (log(+1)-log(−1))
}
}while(s_i ! = NULL);

- Opinions categorizer: Next is to categorize these opinions into positive and negative category by using the Naive Bayes. The well-known probabilistic classifier is Naïve Bayes Classifier. It is used to describe its application to text for incorporating unlabelled data. The learning of the model is conducted by including parameters which are predicted by using labeled training data. In the proposed algorithm, nouns which are taken as features are used to calculate probabilities of the positive and negative count [45].

The algorithm for Naive Bayes Classifier is given in Table 2:

Input: Sentences $\{s_1 + s_2 + s_3 + \ldots s_n\}$ divided into List of words (tokens) words = $\{wo_1 + wo_2 + wo_3 + \ldots wo_n\}$ where x = 1, 2, 3, …, n
Database: Naive Table T_d
Positive words: $\{pwo_1 + pwo_2 + pwo_3 + \ldots p\ wo_n\}$
Negative words: $\{nwo_1 + nwo_2 + nwo_3 + \ldots nwo_n\}$
Neutral words: $\{nuwo_1 + nuwo_2 + nuwo_3 + \ldots nuwo_n\}$
Positive Polarity =+1
Negative Polarity = −1

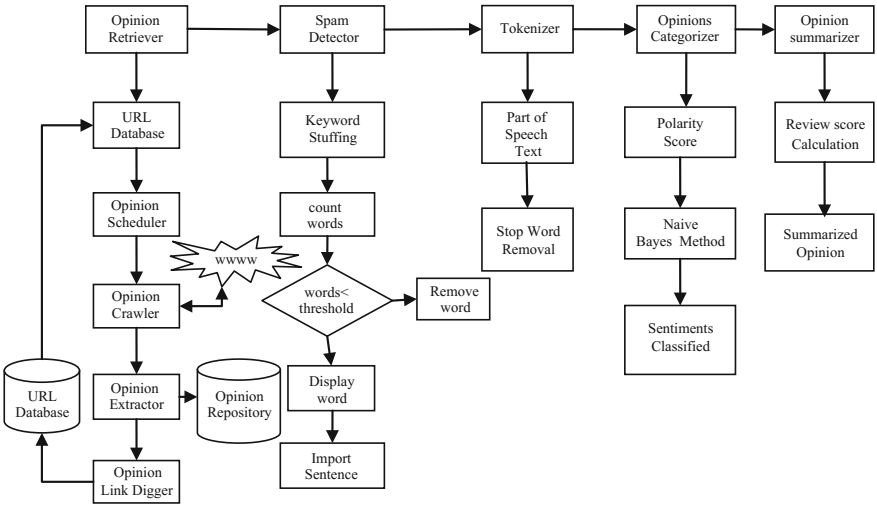

Fig. 5 Detailed architecture

Opinion Summarizer: After analyzing the sentiments and calculating the review score of the positive instances as well as negative instances, the summary is produced as to whether a product is recommended to purchase or not.

The complete model of the proposed work is given in Fig. 5.

4 Results and Analysis

The following section conducts the experiments on the sample data and analysis is done by calculating performance metrics.

4.1 Dataset Description

The customer review dataset of a product is used for our analysis. The tweets are collected from the http://www.Twitter.com. Opinions may contain complete sentences as reviews or may be rated as stars with date and time but tweets generally consists of short comments. LG television product reviews are used in our work. These opinions are categorized into individual sentences. The dataset used in the proposed system is shown in Table 3.

Table 3 Corpus details

S No.	Corpus	LG LED television
1	Opinions	150
2	Total sentences	460
3	Positive sentences	252
4	Negative sentences	143
5	Total opinion as sentences	395
6	Percentage	85.86%

4.2 Evaluation

The three basic measures explained as under [46] are used for evaluation of our proposed framework and comparison is done.

The ratio of relevant tweets retrieved to the total number of tweets retrieved (relevant and irrelevant tweets retrieved is called Precision. Mathematically,

$$\text{Precision} = \frac{\text{RTT}}{\text{RTT} + \text{RWT}} \tag{1}$$

where, RTT is the relevant tweets retrieved and RWT is the irrelevant tweets retrieved.

Recall is defined as the ratio of relevant tweets retrieved to the manually retrieved tweets by the classifier (relevant tweets retrieved and relevant tweets not retrieved). Mathematically,

$$\text{Recall} = \frac{\text{RTT}}{\text{RTT} + \text{RNT}} \tag{2}$$

where, RTT is number of relevant tweets retrieved and RNT are relevant tweets not retrieved.

F-Measure is explained mathematically [47].

$$\text{F-measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

The values are identified manually in the system to calculate the systems accuracy with these performance metrics.

4.3 Result

The test set used is LG Television which is annotated below using positive and negative polarities. On determining the true values, Opinions categorizer provides 92.36% of accuracy for the given dataset. The parameters for opinions categorizer are shown in the Fig. 6.

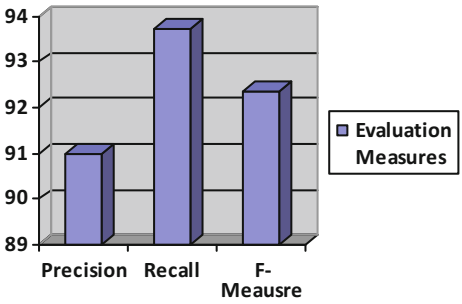

Fig. 6 Parameters for evaluation

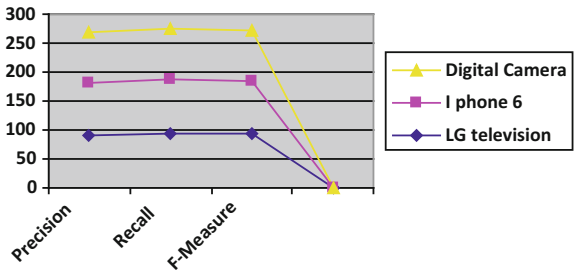

Fig. 7 Different products evaluation

Precision, recall and accuracy rates are calculated on the basis of features per sentence (Fig. 7). We did the same work for electronics and healthcare domain. Tests are done by collecting reviews also for I phone 6, Digital camera Sony and Bausch and Lomb Contact lens which are displayed in Fig. 8. Table 4 lists the details of the results.

The analysis of our work is demonstrated using WEKA[1] tool by using the data set "Bausch and Lomb contact lens.arff". We have taken first twenty four instances for analysis which consists of training and test data. The data consists of 14 positive and 9 negative samples. The snapshot of the Naive Bayes algorithm is shown in Fig. 8.

5 Future Trends and Conclusion

A novel framework is designed that provides relevant features and benefits in recommendation of a product by considering reviews. This chapter presents a research plan with an overarching goal to help ensure that the proposed opinion system is developed systematically with novel algorithms. The result of this plan will be a good approach to improve decision making policies of customers, made available to a wider community of academics and practitioners. The chapter concludes with a summarized opinion, a balanced assessment of the contribution of mining opinions and a roadmap for future directions. There is still a gap needed to

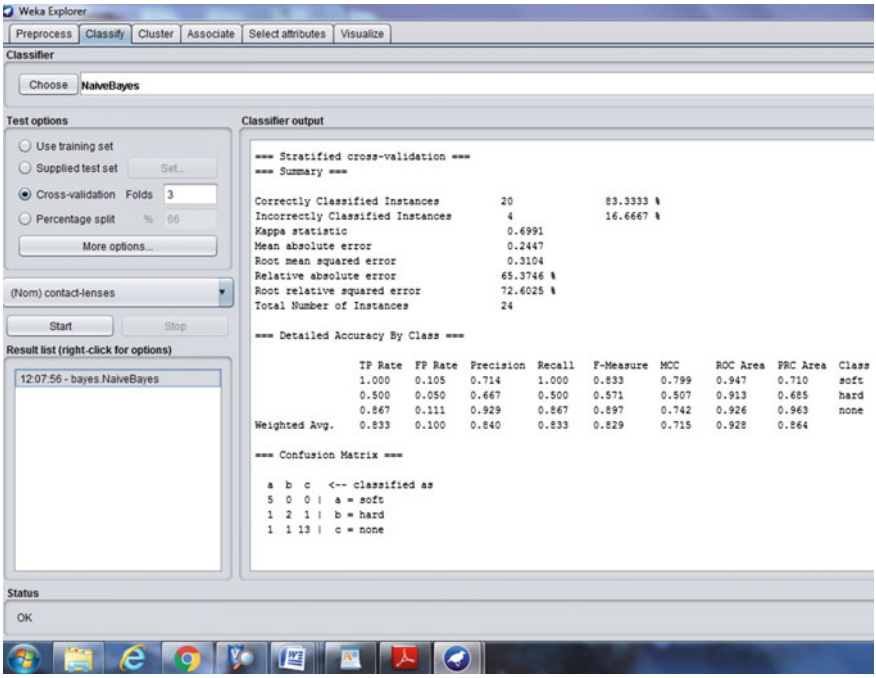

Fig. 8 Naïve Bayes algorithm

Table 4 Results

Product	Precision (%)	Recall (%)	F-measure (%)
LG television	91	93.73	92.36
I Phone 6	90	92.8	91.4
Digital camera sony	86.2	87.5	86.85

be filled to make our system more user friendly as compared with human annotation.

Our future work will include the analysis of various supervised, unsupervised and semi supervised learning techniques in detecting sentiments. The semantic (knowledge) data present in the review needs to be analyzed to detect sentiment by developing the opinion ontology.

5.1 Scope of Chapter

This chapter defines the scope of Internet of Things and big data, as covered in this book, discusses the need of mining opinions, developing information retrieval system, and outlines some of the benefits associated with the social interaction and

social software on World Wide Web. In addition to the theoretical aspects of opinion mining, it covers various supervised learning algorithms used in the classification of opinions, and describes a typical coherent framework of how to mine opinions. The rapid evolution of Electronic media led to the growth of social media users to 127 million users in the year 2013 which will further increase substantially close to 283 million users by 2017. Therefore, the dependency of public on search engines and Information retrieval systems has become amplified. Opinion mining lies in close proximity with social media. As the time changes and E-business takes its place through Social media, demand of a new web world has awaken people to share views and opinions in order to make correct decisions by extracting useful content (aspects) followed with summarization of opinions with their polarity score. Interest of researchers towards the field of Opinion mining is a demanding task. Thus, a true need has been raised to mine opinions and develop an opinion mining system which will help people in taking better decisions of buying and selling products online considering their relevant aspects.

References

1. Bhatia, S., Sharma, M., & Bhatia, K. K. (2015). Strategies for mining opinions: A survey (2nd Edn.). In *International conference on computing for sustainable global development (INDIACom)* (pp 262–266). New Delhi: IEEE Xplore.
2. Abulaish, M., Doja, M. N., & Ahmad, T. (2009). Feature and opinion mining for customer review summarization. In *Pattern recognition and machine intelligence* (pp 219–224). Berlin: Springer.
3. Khan, K., Baharudin, B., Khan, A., & Ullah, A. (2014). Mining opinion components from unstructured reviews: A review. *Journal of King Saud University-Computer and Information Sciences, 26*(3), 258–275.
4. Seerat, B., & Azam, F. (2012). Opinion mining: Issues and challenges (a survey). *International Journal of Computer Applications* (0975–8887)*, 49*(9).
5. Vinodhini, G., & Chandrasekaran, R. M. (2012). Sentiment analysis and opinion mining: a survey. *International Journal, 2*(6).
6. Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies, 5*(1), 1–67.
7. Lo, Y. W., & Potdar, V. (2009). A review of opinion mining and sentiment classification framework in social networks. In *2009 3rd IEEE international conference on digital ecosystems and technologies* (pp. 396–401). New York: IEEE.
8. Khan, F. H., Bashir, S., & Qamar, U. (2014). TOM: Twitter opinion mining framework using hybrid classification scheme. *Decision Support Systems, 31*(57), 245–257.
9. Penalver-Martinez, I., Garcia-Sanchez, F., Valencia-Garcia, R., Rodríguez-García, M. Á., Moreno, V., Fraga, A., et al. (2014). Feature-based opinion mining through ontologies. *Expert Systems with Applications, 41*(13), 5995–6008.
10. Song, Q., Ni, J., & Wang, G. (2013). A fast clustering-based feature subset selection algorithm for high-dimensional data. *IEEE Transactions on Knowledge and Data Engineering, 25*(1), 1–4.
11. Dongre, A. G., Dharurkar, S., Nagarkar, S., Shukla, R., & Pandita, V. (2016). A survey on aspect based opinion mining from product reviews. *International Journal of Innovative Research in Science, Engineering and Technology, 5*(2), 2319–8753.

12. Mfenyana, S. I., Moroosi, N., Thinyane, M., & Scott, S. M. (2013). Development of a facebook crawler for opinion trend monitoring and analysis purposes: case study of government service delivery in Dwesa. *International Journal of Computer Applications, 79* (17).
13. Zhang, S., Jia, W. J., Xia, Y. J., Meng, Y., & Yu, H. (2009). Opinion analysis of product reviews. In *Fuzzy Systems and Knowledge Discovery, 2009. FSKD'09*. Sixth International Conference on 2009 August 14 (Vol. 2, pp. 591–595). New York: IEEE.
14. Varghese, R., & Jayasree, M. (2013). Aspect based sentiment analysis using support vector machine classifier. In *Advances in computing, communications and informatics (ICACCI), international conference on IEEE*.
15. Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-2002 July 6* (Vol. 10, pp. 79–86). Association for Computational Linguistics.
16. Hemalatha, I., Varma, D. G., & Govardhan, A. (2013). Sentiment analysis tool using machine learning algorithms. *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 2*(2), 105–109.
17. McDonald, R., Hannan, K., Neylon, T., Wells, M., & Reynar, J. (2007). Structured models for fine-to-coarse sentiment analysis. In *Annual meeting-association for computational linguistics 2007 June 23* (Vol. 45, No. 1, p. 432).
18. Hu, M., & Liu, B. (2004). Mining opinion features in customer reviews. In *AAAI 2004 July 25* (Vol. 4, No. 4, pp. 755–760).
19. Bhatia, S., Sharma, M., & Bhatia, K. K. (2016). A novel approach for crawling the opinions from world wide web. *International Journal of Information Retrieval Research (IJIRR), 6*(2), 1–23.
20. Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University, 33*(2004), 1–26.
21. Ganesan, K. A., & Kim, H. D. (2008). Opinion mining—A Short Tutorial (Talk). University of Illinois at Urbana Champaign.
22. Sharma, N. R., & Chitre, V. D. (2014). Opinion mining, analysis and its challenges. *International Journal of Innovations & Advancement in Computer Science., 3*(1), 59–65.
23. Ma, Z. M. (2005). Databases modeling of engineering information. China: Northeastern University.
24. Seerat, B., & Azam, F. (2012). Opinion mining: Issues and challenges (a survey). *International Journal of Computer Applications, 49*(9).
25. Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of annual meeting of the association for computational linguistics* (ACL-2002).
26. Wiebe, J., Bruce, R. F., & O'Hara, T. P. (1999). Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the association for computational linguistics* (ACL-1999).
27. Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of ACM SIGKDD international conference on knowledge discovery and data mining* (KDD-2004).
28. Jedrzejewski, K., & Morzy, M. (2011). Opinion mining and social networks: A promising match. In *Advances in social networks analysis and mining (ASONAM), 2011 international conference on 2011 July 25* (pp. 599–604). New York: IEEE.
29. Kasthuri, S., Jayasimman, L., & Jebaseeli, A. N. An opinion mining and sentiment analysis techniques: A survey. *International Research Journal of Engineering and Technology (IRJET), 3*(2). e-ISSN: 2395–0056.
30. Agrawal, R., Rajagopalan, S., Srikant, R., & Xu, Y. Mining newsgroups using networks arising from social behavior. In *Proceedings of the 12th international conference on world wide web 2003 May 20* (pp. 529–535). ACM.

31. Vijaya, D. M., & Sudha, V. P. (2013). Research directions in social network mining with empirical study on opinion mining. *CSI Communication, 37*(9), 23–26.
32. Stavrianou, A., Velcin, J., & Chauchat, J. H. (2009). A combination of opinion mining and social network techniques for discussion analysis. *Revue des Nouvelles Technologies de l'Information*. 25–44.
33. Buche, A., Chandak, D., & Zadgaonkar, A. (2013). Opinion mining and analysis: A survey. arXiv preprint arXiv:1307.3336
34. Sobkowicz, P., Kaschesky, M., & Bouchard, G. (2012). Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web. *Government Information Quarterly, 29*(4), 470–479.
35. Gao, Wei, & Sebastiani, Fabrizio. (2016). From classification to quantification in tweet sentiment analysis. *Social Network Analysis and Mining, 6*(19), 1–22.
36. Montoyo, A., MartíNez-Barco, P., & Balahur, A. (2012). Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. *Decision Support Systems, 53*(4), 675–679.
37. Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems, 28*(2), 15–21.
38. Osimo, D., & Mureddu, F. (2012). *Research challenge on opinion mining and sentiment analysis*. Universite de Paris-Sud, Laboratoire LIMSI-CNRS, Bâtiment, p 508.
39. Xia, R., Zong, C., & Li, S. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences, 181*(6), 1138–1152.
40. Melville, P., Gryc, W., & Lawrence, R. D. (2009). Sentiment analysis of blogs by combining lexical knowledge with text classification. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining 2009 June 28* (pp. 1275–1284). ACM.
41. Zhang, Z., Ye, Q., Zhang, Z., & Li, Y. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications, 38*(6), 7674–7682.
42. Tan, S., & Zhang, J. (2008). An empirical study of sentiment analysis for chinese documents. *Expert Systems with Applications, 34*(4), 2622–2629.
43. Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications, 36*(3), 6527–6535.
44. Tretyakov, K. (2004). Machine learning techniques in spam filtering. In *Data mining problem-oriented seminar, MTAT 2004 May 3* (Vol. 3, No. 177, pp. 60–79).
45. Bhatia, S., Sharma, M., & Bhatia, K. K. (2015). Sentiment knowledge discovery using machine learning algorithms. *Journal of Network Communications and Emerging Technologies (JNCET), 5*(2), 8–12.
46. Hildreth, C. R. (2001). Accounting for users' inflated assessments of on-line catalogue search performance and usefulness: An experimental study. *Information Research, 6*(2), 6–2.
47. Jeyapriya, A., & Selvi, C. K. (2015). Extracting aspects and mining opinions in product reviews using supervised learning algorithm. In *Electronics and communication systems (ICECS), 2015 2nd international conference on 2015 February 26* (pp. 548–552). New York: IEEE.