SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM

TMLAI TRANSACTIONS ON
MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Volume 6 No. 1
ISSN 2054-7390

Extracting Sentiments and Summarizing Health Reviews from Social Media Using Machine Learning Techniques

[1]Mozibur Raheman Khan , [2]Rajkumar Kannan
Department of Computer Science, Bishop Heber College (Autonomous),
Tiruchirappalli, India.
[1]mozibmsc@gmail.com, [2]rajkumar@bhc.edu.in

ABSTRACT

Most of the health organizations provide an array of medical services and request their beneficiaries to provide their experience's in the form of opinion/reviews for which they are associated. Doctors of national and international repute have hundreds and even thousands of reviews authored by the health consumers around the globe. For an individual it is difficult and time consuming process to look all the reviews before taking an appropriate decision. Thus it is necessary to summarize the reviews to make an individual to take prompt decision. For a doctor it is also difficult to keep track of patient's reviews given by the patients in different time intervals, but he may have the summary of his entire patient's reviews to understand what is the best can be done to the patient's community. This research paper aims to mine and summarize the medical reviews authored by the health consumers. This article is performed by summarization of text in three steps, the first step is to identify the health features that have been commented by health consumers, the next one is to identify opinions of each review sentence and deciding whether each opinion sentence is positive or negative and finally summarizing the results.

Keywords—national and international repute, medical reviews, health consumer, summarization of text and health features

1 Introduction

The growth and expansion of internet, more and more services are provided on the Web, and more and more people are also deriving the benefits of the offered services. To provide better online services and to make prompt decision it has become a common practice for online health service provider to enable their health consumer to provide reviews or to express their opinions about the services they are enjoying. Reputed service organization gets large number of reviews from across the different region by different people, for an individual it is time consuming process to read the entire reviews.

In some cases reviews may be long and in some cases opinions are reflected for a particular feature. If a person wants to visit a hospital to access the health related services, he/she may not take an appropriate decision by reading few reviews. Even if he/she takes the decision, the decision may be biased. The suffering of public and in general health consumers due to poor medical facilities and less expertise of health consultant has increased their suffering by many folds. Hence it is advised to look the summary of the large reviews before making a final decision. Thus the need arises to collect the reviews that express

DOI: 10.14738/tmlai.61.3595
Publication Date: 16th, December 2017
URL: http://dx.doi.org/10.14738/tmlai.61.3595

their views for particular services they are willing to access or before willing to buy any product. Then these reviews are summarized to help the health consumers to approach the appropriate medical consultant. 'Opinions' mainly include opinionated text data such as blog/review articles, and associated numerical data like aspect rating is also included [1].some of the websites provides the very useful information related to health domain. Figure1 is screen shot taken from www.ratemds.com which provides numerical ratings as well as corresponding reviews of the different health experts.

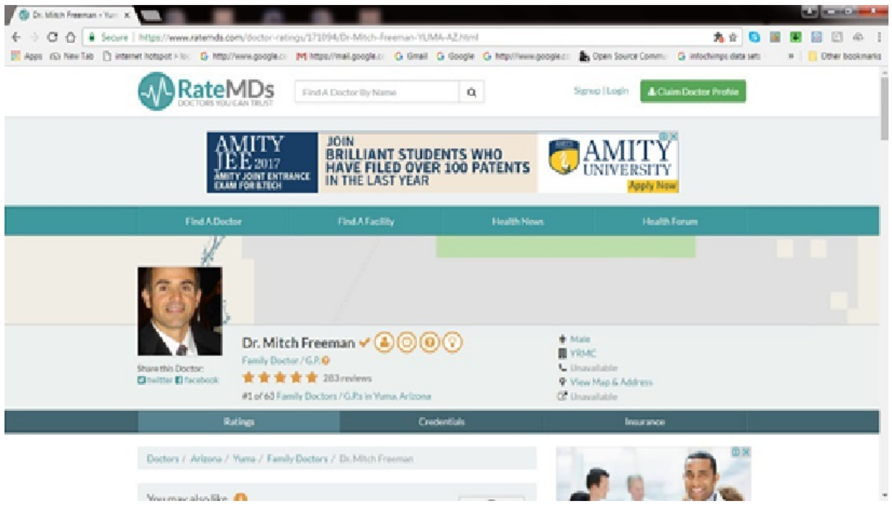


Figure 1: Screen shot with its reviews obtained from www.ratemds.com

The focus of this research is to study the problem of generating *feature-based summaries* of medical reviews given by different health consumers at various time intervals. Here, *features* broadly mean the doctor features (or attributes) and the attribute of its supporting teams members.

Given a set of medical reviews of a particular health consultant, the task involves three subtasks, the first step is identifying health features of the doctor that health consumers have expressed their opinions, the next for each feature, identifying review sentences that give positive or negative opinions; and finally producing a summary using the discovered information. Below is an example of feature-based summary. Consider the reviews of a particular doctor say, health expert. The summary looks like the following:

Summary of a Health Expert

Feature: staff
Positive:
Dr. Freeman's staff is very friendly and his knowledge alone is worth the extra money. I have always left with a good experience and highly recommend him to others looking for a doctor in YumaSubmitted Oct. 27, 2015
Negative:

Feature: Punctuality
Positive:
He is good
Negative:

Feature: Recommend
Positive: "Dope doctor! I recommend him. And i don't usually do this but i mean it. 👍"
Negative:

Feature: Knowledge
Positive:
Dr. Freeman's staff is very friendly and his knowledge alone is worth the extra money. I have always left with a good experience and highly recommend him to others looking for a doctor in YumaSubmitted Oct. 27, 2015
Best Dr. I've ever seen. He is Knowledgeable and cares.
Negative:
This Dr. Dose not except Medicare without charging 100$ a month extra

Figure 2: An example of summary

In Figure 2, staff and Punctuality are the doctor features. We have one medical review that express positive opinions about the staff, and knowledge and one review that express negative opinions for knowledge. With such a feature-based summary, one may understand the general opinion for a particular doctor. If he/she is very interested in a particular feature, he/she can drill down by following the individual review sentences to understand the level of satisfaction of health consumer or what may be the complaint. For a doctor of high repute/hospital of high repute may look the summary to understand what actually they are doing? And what supposed to be done, so that they can provide the services to suite the requirement of the health consumer.

Our task is different from traditional text summarization [9-11] in a number of ways. This health review summary in our case is *structured* rather than another (but shorter) free text document as produced by most text summarization systems. Second, we are only interested in features of the doctor that patients have opinions on and also whether the opinions are positive or negative. Traditional text summarization captures all the original text and important points but we follow the different techniques to summarize the health reviews.

As indicated above, our task is performed in three main steps; the first step is to capture health features that has been commented by patients. Data mining and natural language processing techniques are used to perform our task. This part of the study has been reported in [19].However, for completeness, we will summarize its techniques in this paper and also present a comparative evaluation.

The next step is to Select the reviews consists of opinion sentences and determine whether each opinion sentence is positive or negative. Note that these opinion sentences must contain one or more health features identified above. *Opinion orientation* of each sentence is determined (whether the opinion expressed in the sentence is positive or negative), by performing three subtasks. First, a set of adjective words (which are normally used to express opinions) is identified using a natural language processing method. For the selected features we have corresponding opinion and these opinions are called as *opinion words.*

Summarizing the results. This step aggregates the results of previous steps and presents them in the format of Figure 2.Section 3 presents the detailed techniques for performing these tasks. A system, called Health Review Summarization has also been implemented. Our experimental results with a large number of medical reviews of doctor available online show that health review summarization system (HRS) and its techniques are highly effectiveness.

Rest of the paper is organized as follows. In Section 2, related works are presented. In Section 3, feature Based Opinion Summarization approach is introduced. In Section 4, experiments and results are presented. In Section 5, the conclusion is presented.

2 Related Work

This work is related to on Mining and Summarizing Customer Reviews [2]. The system performs the summarization in three main steps (as discussed before),the first step is mining product features that have been commented on by customers, the next one is identifying opinion sentences in each review and deciding whether each opinion sentence is positive or negative, and finally summarizing the results. These steps are performed in multiple sub-steps.

Given the inputs, the system first downloads all the reviews, and put them in the review database. We then look very “hot” (or frequent) features that most of the people have expressed their opinions on. Collect the opinion words and determined its semantic orientations of the opinion words. Once opinion words are extracted, then the system then finds those infrequent features. In the last two steps, they found the orientation of each opinion sentence is identified and a final summary is produced. Note that POS tagging [28] is used from natural language processing, which helps us to find the features.

Another work is related to on semantic classification of reviews [3]. Using available training corpus from some Web sites, where each reviews already has a class (e.g., thumbs-up and thumbs-downs, or some other quantitative or binary ratings), they designed and sentiment classifier is built after experimenting a number of methods. They have shown that such classifiers perform quite well with test reviews and classifiers is used to classify sentences obtained from Web search results, which are obtained by a search engine using a product name as the search query.

Collecting individual sentences/opinion from the web searches, performance is limited due to noise and ambiguity. But in the context of a complete web-based tool and aided by a simple method for grouping sentences into attributes, the results are qualitatively quite useful.

The reputation of the target product by compare reviews of different products in one category is discussed in [4]. However, it does not summarize reviews, and it does not mine product features on which the reviewers have expressed their opinions. Although they do find some frequent phrases indicating reputations, these phrases may not be product features (e.g., “doesn’t work”, “benchmark result” and “no problem(s)”).Knowing the reputations of your own and/or competitors' products is important for marketing and customer relationship management. It is, however, very expensive to collect and analyze survey data manually. This paper express the reputations of doctors globally on the Internet and health reviews are downloaded or can be crawled automatically those express health consumer opinions for the concerned expert or the health service provider.

In [5], the discussion on opinion-oriented information extraction. Their aim is to create summary representations of opinions to perform question answering.They propose to use of opinion-oriented "scenario templates" to act as summary representations of the opinions expressed in a document or a set of documents.

Our approach is different. We initially interested to identify the doctor features and user opinions on these features to automatically produce a summary. This work is also partially related but different from subjective genre classification, sentiment classification, text summarization and finding the terminology. It is discussed by each of them below.

2.1 Subjective Genre Classification

Genre classification classifies texts into different styles, e.g., "editorial", "novel", "news", "poem" etc. Although some techniques for genre classification can recognize documents that express opinions [6-8], they do not tell whether the opinions are positive or negative. In this work, we have to determine the opinion polarity and to perform opinion classification at the sentence level rather than at the document level.

A more closely related work is [12], in which the authors investigate sentence subjectivity classification and concludes that the presence and type of adjectives in a sentence is indicative of whether the sentence is subjective or objective. However, their work does not relate to our task of determining the semantic orientations of those subjective sentences. Even they neither find the features nor interested on which features opinions have been expressed.

2.2 Sentiment Classification

The phrase *sentiment analysis* is closely resembles with that of "opinion mining" in certain respects**.** The term "sentiment" is used in reference to the automatic analysis of evaluative text and tracking of the predictive judgments there in appears in 2001 papers [13, 14] because of these authors' interest in analyzing market sentiment. They use a manually crafted lexicon in conjunction with several scoring methods to classify stock postings on an investor bulletin. It subsequently occurred within 2002 papers [32,33] which were published in the proceedings of the annual meeting of the Association for Computational Linguistics (ACL) and the annual conference on Empirical Methods in Natural Language Processing (EMNLP).

Works of [15,16] on sentiment-based classification of entire documents use models inspired by cognitive linguistics. The author work in [43] also manually constructs a discriminant-word lexicon and use fuzzy logic to classify sentiments. Generates sentiment timelines is reported in [14]. The author captures online discussions about movies and displays a plot shown with the number of positive and negative sentiment messages over time. Messages are classified by looking for specific phrases that indicate the author's sentiment towards the movie (e.g., "great acting", "wonderful visuals", "uneven editing"). Each phrase are taken manually and added to a special lexicon and manually tagged as indicating positive or negative sentiment. The lexicon is domain dependent (e.g., movies) and must be rebuilt for each new domain. In our work, this research is concerned with most frequent features and applied classification techniques.

Applies a specific unsupervised learning technique based on the mutual information between document phrases and the words "excellent" and "poor", where the mutual information is computed using statistics gathered by a search engine[17].Examine several supervised machine learning methods for sentiment classification of movie reviews and conclude that machine learning techniques performs well with other method that is based on human-tagged features although none of existing methods could handle the sentiment classification with a reasonable accuracy[18].

A sizeable number of papers mentioning "sentiment analysis" focus on the specific application of classifying reviews as to their polarity (either positive or negative), a fact that appears to have caused some authors to suggest that the phrase refers specifically to this narrowly defined task. However, the term "sentiment analysis" more broadly used by researchers to mean the computational analysis of opinion, sentiment, and subjectivity in text**.** "Sentiment analysis" and "opinion mining" denote the same field of study (which itself can be considered a sub-area of subjectivity analysis).They have attempted to use these terms more or less interchangeably in their work.

A novel approach is proposed based on latent semantic analysis (LSA) to identify product features [46]. Furthermore, they found a way to reduce the size of summary based on the product features obtained from LSA. They consider both sentiment-classification accuracy and system response time to design the system. This system can be extended to other product-review domains easily. This research paper is mainly based on movie reviews collected from Internet Blogs that do not consist of any rating information. Sentiment analysis is performed to determine the semantic orientation of the reviews and movie-rating score is based on the sentiment-analysis result. In addition to the accuracy of the classification, system response time is also taken into account in our system design. Although they have focused on movie review, the whole design is not only for movie-review domain. They have performed sentiment classification on movie review dataset, which is available[1] .The dataset includes 1000 positive and 1000 negative movie reviews. Similarly, SVM is used to perform the classification task. The kernel function used in the system is RBF and *K*-fold cross validation (i.e., $K = 5$) is used in the experiment.

2.3 Text Summarization

Summarization technique can be of two types. The first one Extractive Summary is a summary that represent by selecting representative text segments, usually sentences, from the original documents and another one is Abstractive Summary does not use the existing sentences in representing the summary rather it analyzes documents and directly generates sentences.

[1]http://www.cs.cornell.edu/People/pabo/movie-review-data

Because it is very tedious to produce readable and complete sentences, studies on extractive summary are more popular than that on abstractive summary. Extracting salient sentences from text and coherently organizing them to build a summary of the entire text is the key area of summarizing documents that focused on proposing paradigms**.** The relevant works in this regard includes [27, 28 and 47]. While traditional works focused on summarizing a single document, later, researchers shifted the idea on summarizing multiple documents originated from multiple sources.

The definition of a summary as a text that is generated from one or more texts, that conveys what the original text conveys, and that is lesser than that of the actual text(s) and must be less than that of [53]. This simple definition provides three aspects of automatic text summarization:

Single document or multiple documents can be summarized

Summaries should preserve the essence of the original text or paragraph

The length of the Summaries should be short.

Even if we agree unanimously on these points, it seems from the literature that any attempt to provide a more elaborate definition for the task would result in disagreement within the community. In fact, many approaches differ on the manner of their problem formulations. Some common terms introduced in the summarization dialect: extraction is the procedure of identifying important sections of the text and producing them verbatim; abstraction aims to produce important material in a new way; fusion makes an attempt to combines extracted parts coherently; and compression aims to throw out unimportant sections of the text [53].

The authors in [23, 24] emphasize on identification and extraction of certain core entities and facts in a document, which are packaged in a template. This framework requires background knowledge in order to instantiate a template to a suitable level of detail. Therefore, it is not domain or genre independent [25, 26]. This is different from our work as our techniques do not fill any template and are domain independent. The passage extraction framework [e.g., 27-29] identifies certain segments of the text (typically sentences) that are the most representative of the document's content. Our work is different in that we do not extract representative sentences, but identify and extract those specific product features 1 and the opinions related to them. An idea is proposed to find a few very prominent expressions, objects or events in a document and use them to help summarize the document proposed [31]. This work is again different as we find all health related features from a set of health consumer review regardless whether they are prominent or not. Thus, our summary is not a traditional text summary.

Lots of works have been done on text summarization focusing on a single document. Recently few of researchers also studied on summarization of multiple documents covering similar information. The authors in[27] have Summarized the similarities and differences in the information content is the focus of their work**.**Our work is related but quite different because we take interest to find the key features that are discussed by multiple reviews. Summarizing the similarities and differences of reviews is not the key focus.

Opinion summarization has different aspects from the classic text summarization problem because the nature and structure of the data**.** While summarizing the opinion, usually the polarities of input opinions are crucial. Sometimes, those reviews are provided with additional information such as rating scores. The formats of the summary is proposed by the most of the researcher of the opinion summarization are more structured in nature with the segmentation by topics and polarities. However, techniques of text summarization still can be useful in opinion summarization when text selection and generation step. Once separating input data by its polarities and topics, classic text summarization techniques can be used to find/generate the most representative text snippet from each category.

In health-review summarization, generally health consumer is more interested to know the expertise of a particular doctor in a specific field. To understand the expertise of a doctor, we must look what opinion the expert receives for a specific feature from the reliable/trusted sources. Hence it is necessary to understand the important features of doctor and the corresponding opinion. Thus, feature-based summarization is used in health-review summarization. The feature-based summarization will focus on the doctor features on which the patients or public have expressed their opinions. In addition to doctor features, the summarization should include opinion information about the doctor or concerned hospitals; therefore, doctor features and opinion words are both important in feature-based summarization. As a result, doctor's features and opinion-word Identification are essential in feature-based summarization.

In case of feature-based summarization we are very much interested to find out the aspects and these salient aspects is given as an input, which is also called as features and subtopics, and generates summaries of each feature. For example, for the summary of 'doctor', there can be aspects such as 'punctuality', 'knowledge', 'care', 'cost', etc. By further splitting the input texts into smaller units, aspect-based summarization can show more details in a structured way. Further splitting of feature can be even more useful when overall opinions are different from opinions of each aspect because aspect-based summary distribute the opinion of each aspect separately. The feature-based approaches are very popular and have been heavily explored over the last few years [44].

2.4 Summary Generation

Using the results of feature discovery and sentiment prediction, it is then critical to generate and present the final opinion summaries in an effective and easy to understand format. This typically involves aggregating the results of the first two steps and generating a concise summary. The following techniques describe various generation methods for opinion summarization**.** Each technique has its own advantages and disadvantages and some techniques can be combined with others. For example, we may add a timeline to text selection methods**.**

Statistical Summary. The most popular format and commonly adopted is a summary showing statistics introduced in [44].Statistical summary directly uses the processed results from the previous two steps - a list of aspects and results of sentiment prediction. All positive and negative opinions for each aspect can be displayed, so that the readers can easily understand the overall sentiments of users at large. Along with the positive and negative occurrences, all sentences with sentiment prediction in each aspect is shown (Figure 1).

The author has showed statistics in a graph format [49]. With the graph based representation, they collect people's overall opinions about the target more intuitively. Opinion observer is software developed by Liu et al. in 2005 clearly shows the statistics of opinion orientation in each aspect and it allows the users to compare opinion statistics of several products. An example result is shown in Figure 2, which gives the summary of different doctor.This format of summary has been widely adopted even in the commercial world.

Text Selection. While statistical summaries help users understand the overall idea of people's opinion, sometimes reading actual text is necessary to understand specifics. Due to the large volume of opinions on one topic, showing a complete list of sentences is not very useful.

Aggregated Ratings. Proposed the advanced summary is reported in [50], *aggregated ratings*, which combine statistical summary and text selection. Based on the discovered aspects using clustering and topic modeling, they average the sentiment prediction results of phrases for each aspect as the final sentiment rating for that aspect. Aspect ratings are shown with representative phrases.

Summary with a Timeline. Opinion trends over a timeline reflected in [51, 52]. General opinion summarization focuses on finding statistics of the 'current' data. In reality, opinions change as time goes by. Opinion summary with a timeline helps us to see the trend of opinions on the target easily, and it also tells ideas to further analysis. To figure out what changes people's opinions, we can analyze the events that happened at the drastic opinion change.

2.5 Terminology Finding

In terminology finding, there are basically two techniques for is discovering terms in corpora: symbolic approaches that rely on syntactic description of terms, namely noun phrases, and statistical approaches that exploit the fact that the words composing a term tend to be found close to each other and reoccurring [19-22]. However, using noun phrases tends to produce too many non-terms (low precision), while using reoccurring phrases misses many low frequency terms, terms with variations, and terms with only one word.

3 Feature Based Reviews Summarization

Figure3 provides the architectural overview of our health reviews summarization system. The inputs for the system are, a doctor's name and the salient features from the corresponding reviews. The output is the summary of the reviews as the one shown in the introduction section. The system performs the summarization in three main steps (as discussed before), the first step is Mining health features features that have been commented on by health consumers; the second is identifying opinion sentences in each review and deciding whether each opinion sentence is positive or negative, and finally Summarizing the results. These steps are performed in multiple sub-steps**.**

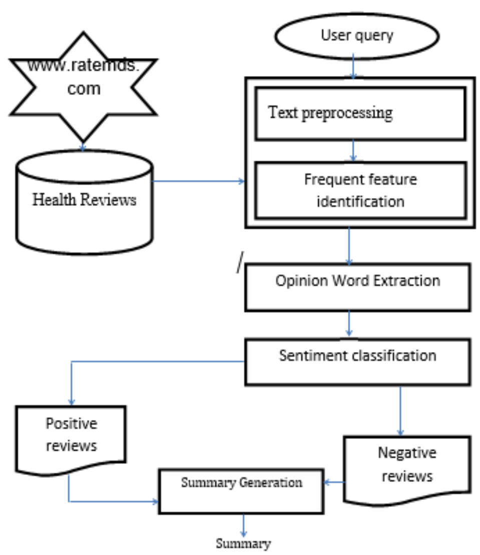


Figure 3: Architecture of Feature-based Reviews Summarization System

3.1 Data Collection and Preprocessing

We have collected a corpus of reviews from www.**ratemds.com** websites and these downloaded reviews were placed in the review database. As a preprocessing step the portions containing the reviews were extracted from html pages and these reviews were tokenized and separated into individual sentences. We then find those frequent features that many health consumers have expressed their opinions on. In the last two steps, the orientation of each opinion sentence is identified and a final summary is produced. Note that POS tagging is the part-of-speech tagging [28] from natural language processing, which helps us to find features and opinion. Below, we discuss each of the sub-steps in turn. Here is the review from the ratemds.com

3.2 Part-of-Speech Tagging (POS)

Identifying the interesting features from the health review are usually nouns or noun phrases. Thus the part-of-speech tagging is crucial. The process also identifies simple noun and verb groups (syntactic chunking). The following figure shows a sentence with POS tags shown in fIgure4.

<S> <NG><W C='PRP' L='SS' T='w' S='Y'> ***I*** </W> </NG>
<VG> <W C='VBP'> ***am*** </W><W C='RB'> ***absolutely***
</W></VG> <W C='IN'> ***in*** </W> <NG> <W C='NN'> ***awe***
</W> </NG> <W C='IN'> ***of*** </W> <NG> <W C='DT'> ***this***

Figure 4: POS tagging

3.3 Frequent Features Identification

Before identification of an individuals persons features on which many people and patients have expressed their opinions, lets us explain what patients like to provide their opinion for respective features before discussing frequent feature identifications. Here is an example of a medical review written by the patients

Great Doctor! For the one of the few times ever, I can say I felt like a doctor was genuinely interested in listening to me and trying to solve my problems!

Figure 5: An example of review authored by health consumer

This sentence expresses the satisfaction of user with the standard service which is provided by a medical consultant. Here the patient's talks about the doctor attribute such as listening and solving the patient's problems. Sometime some features are implicit and hard to find. For example *i like this office because I ve got good service and great Dr.*

Here, the health consumer is talking about the staff *of* the hospital and other feature, but the word staff does not appear in the sentence. In this work, we focus on finding features that appear explicitly as nouns or noun phrases in the reviews. Here, we find the frequent features, i.e., those features that are talked by many health consumers. In our context, an item set is simply a set of words or a phrase that occurs together in some sentences.

The main reason for them to use association mining is because of the following observation. It is common that a customer review contains many things that are not directly related to product features. Different customers usually have different stories. Thus using association mining to find frequent item sets, is appropriate because those frequent item sets are likely to be the product features. When no one talks for a product or product feature, those noun/noun phrases are said to be infrequent are likely to be non-product features.

3.4 Opinion Words Extraction

We now identify opinion words and these words are primarily used to express subjective opinions. Clearly, this is related to existing work on distinguishing sentences used to express subjective opinions from sentences used to objectively describe some factual information [36]. Subjective and objective categories are potentially important for many text processing applications and Work on subjective opinion [37, 38] has established a positive significance correlation with the presence of adjectives. Thus the presence of adjectives is useful for predicting whether a sentence is subjective, i.e., expressing an opinion. This paper uses adjectives as opinion words. Opinion words extraction for those sentences that contain one or more health features, as we are only interested in health consumer's opinions on these health providers**.** Let us first define an opinion sentence.

Definition: *opinion sentence*
If a sentence contains one or more product features and one or more opinion words, then the sentence is called an *opinion sentence.* We extract opinion words in the following manner (Figure 6)

for each sentence in the review database
if (it contains a frequent feature, extract all the adjective Words as opinion words)
nearby adjective is recorded as its effective opinion for each feature in the sentence .
for each feature in the sentence
the nearby adjective is recorded as its *effective opinion*.
/ A nearby adjective refers to the adjacent adjective that modifies the noun/noun phrase that is a frequent feature. */*

Figure 6: Opinion word extraction

3.5 Infrequent Feature Identification

Finding Frequent features are very easy that people normally exchange their comment for given entity. However, there are some features that only a small number of people talked about. These features can also be interesting to some patients/persons willing to derive health benefits and also to the service providers. The question is how to extract these infrequent features (association mining is unable to identify such features)? Considering the following sentences:

"The facility of the hospital is good."

"The location of the hospital is good."

for each sentence in the review database
if (it has no frequent feature but one or more opinion words)
{
Find the nearest noun/noun phrase around the opinion word. These nearest
Noun/noun phrases is said to be infrequent feature.
}

Figure 7: Infrequent feature extraction

Most of the time the nearest noun/noun phrase modifies opinion word**.** This simple heuristic seems to work well in practice. A simple problem exit with the infrequent feature identification using opinion words is that it could find some feature that are irrelevant. There is the reason to use common adjectives to describe a lot of objects, including both interesting features that we want and irrelevant ones. This is not a serious problem because the number of infrequent features, compared with the number of frequent features, is small. They account for around 15-20% of the total number of features as obtained in our experimental results. Infrequent features are generated for completeness. Frequent features are more important than infrequent ones because we need to display the summary of the frequent feature first and then low ranked feature and thus will not affect most of the users.

3.6 Sentiment Classification

Sentiment classification is similar to traditional binary-classification problem. There are many classification techniques are exit for different domains. We used three classification techniques namely Logistic Regression (LR), Support Vector Machine (SVM) and Gaussian Naive Bayes (GNV). Logistic Regression widely used in disciplines ranging from credit and finance to medicine to criminology and other social sciences. Logistic regression is considered to be very effective.

The second one is SVM is a supervised machine learning algorithm which works well with exiting text categorization[46].The goal of this machine learning algorithm is to find a decision boundary between two classes that is maximally far from any point in the training data. There is one interesting property of SVM is that their ability to learn can be independent of the dimensionality of the feature space. The third one we have used is Gaussian Naive Bayes classifiers. All these techniques are not involved in finding the features that is commented by different user.

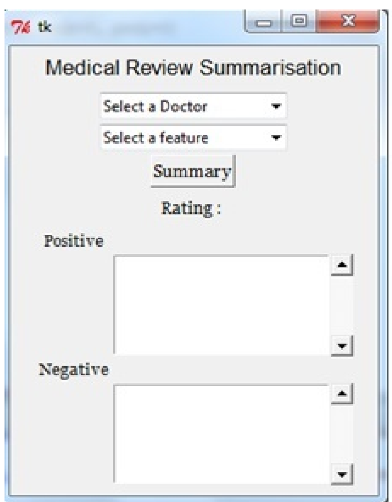


Figure 8: Summarization screen shot

Figure 9: Screen shots with reviews

Figure 8 shows the empty screen receives an input that is a doctor name and the corresponding features then rating is calculated and Figure 9 explains rating and summarizing the particular doctor reviews and we can read all the reviews.

It is easier to some type of probability models that naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods. With small number of training data we can estimate the parameters necessary for classification.

3.7 Summary Generation

Now we are ready to generate the final feature-based review summary, which is straightforward and consists of the following steps:

• For each discovered feature, related opinion sentences are put into positive and negative categories according to the opinion sentences' orientations.

• Ranking of all features is done according to the frequency of their appearances in the reviews. Feature phrases appear before single word features as phrases normally are more interesting to users. Different types of rankings are also possible. For an example, we can choose the rank of features based on the number of reviews that express positive or negative opinions.

The following shows an example summary for the feature "Recommend" of a doctor. It is not necessary that the individual opinion sentences (and their corresponding reviews, which are not shown here) can be hidden using a hyperlink to enable the user to have a quick look of global view of the summary.

Dr. Gretchen
Feature Recommend
Rating *

Positive:
<We highly recommend him>
<Dr. Liddell is wonderful and I recommend him highly to my friends and family>
<She even remembers past conservations we've had! Appointments are readily available but am sure once word gets out how good she is, it will get harder! Highly recommend Dr. Bortolotti.>
<I finally found my Doctor! Took 20 years!!!!She never rushes you out of her office, and if you call to speak to her, SHE calls you back. (instead of a nurse) I would highly recommend!!I highly recommend her. I highly recommend her to everyone.>
....
Negative :

<The only complain is long wait to see her.>
<With that said I highly recommend her...She doesn't just go "by the book.>
<" I highly recommend her!I really like him and his staff, but have had some trouble with getting prescriptions filled in a timely manner, which I found frustrating, but was only an issue because I was in and out of town (and may have been>
...

Figure 10: Review summarization of a health service provider

4 Experiments and results

4.1 Data Sets

We have collected health reviews of one fifty doctor from ratemds.com and these collected reviews have been placed in reviews database. This site provides hundreds of reviews for thousands of doctor from across the globe. Each of the reviews includes a text review and other numeric ratings are available for various other features. We have received all these from family doctor/GP. The site provides numerical rating of four aspects namely staff, punctuality, helpfulness and knowledge. Textual comments are written by the health consumers with an average of three sentences. For each doctor, we first downloaded the first available reviews. Looking at the sites nearly we can understand that there are ten important specialty available. They are Internist, Gynecologist,Family/general, peddiatrist ,Dentist,Psychiatrist, Orthopedist,Cardiologist, Gastroenterologist, Dermatologist and so on. For each specialty there are top reputed doctors are available and each doctor is receiving hundreds of reviews. For hundred and fifty doctor, we have collected 1745 reviews and these reviews are summarized.

Figure 11: Screen shot of health review for a doctor Mitch Freeman

This proposed technique has been implemented in python and we now evaluate our sentiment's extraction and health reviews summarization system from the classification perspectives. We have a corpus of 1745 reviews and these review documents were then cleaned to remove HTML tags. After that, NLP preprocessing techniques is used to generate part of-speech tags. Our system is then applied to perform summarization.

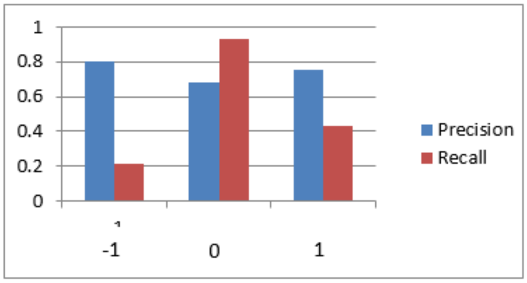
We must identify the orientation of the opinion, is positive or negative. If the user gives no opinion in a sentence, the sentence is not tagged as we are only interested in sentences with opinions in this work. There is a small complication in feature tagging is that features can be explicit or implicit in a sentence. Most features appear explicitly in opinion sentences, e.g., *punctuality in "The wait is a little long but well worth the time spent".* Some features may not appear in sentences. We call such features implicit features, e.g., *punctuality in "The doctor manages his appointment time properly".* Both explicit and implicit features are easy to identify by the human tagger.

Another issue is that judging for evaluation, we manually read all the reviews. For each sentence in a review, if it shows user's opinions, all the features on which the reviewer has expressed his/her opinion are tagged and the opinions in reviews can be somewhat subjective. It is not difficult to judge the opinion is whether it is positive or negative, even in a sentence which expresses its opinion clearly. However, deciding whether a sentence offers an opinion or not can be debatable. For some extreme cases, we reached a consensus between the primary human tagger (the first author of the paper) and the secondary tagger (the second author of the paper).

4.2 Results and Discussion

Bar chart provides the precision and recall results of the feature generation function of Feature Based Summarization. We evaluated the results at each step of our algorithm. The figure10, 11 and 12 gives the recall and precision of frequent feature generation for each doctor using three different classifier. The results indicate that the frequent features contain a lot of errors especially in Gaussian Naïve Bayes, i.e., low precision and moderate recall.

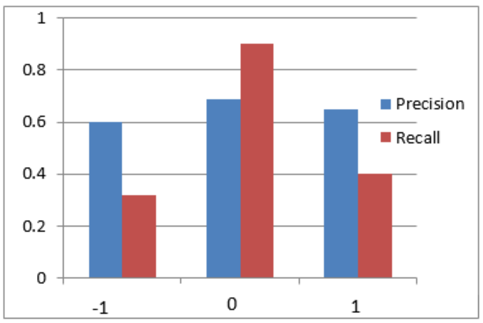
The results can be improved by applying SVM which shows better results than previous one. We can see that the precision is improved marginally and recall is improved drastically. There is another dramatic improvement in the precision by applying logistic regression techniques. The recall level almost does not change comparing to the previous step.



-1 stands for negative sentiments
0 stands for neutral sentiments
1 stands for positive sentiments

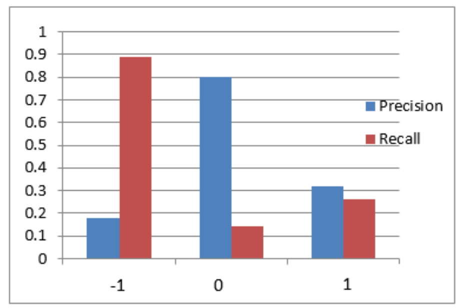
Sentiments

Figure 12: Precision and Recall using LR for Medical reviews data sets



Sentiments

Figure 13: Precision and Recall using SVM for medical reviews data sets



Sentiments

Figure 14: Precision and Recall using GNV for Medical Reviews data sets

5 Conclusion

In this paper, we proposed a set of techniques for extracting sentiments and summarizing the health reviews based on data mining and natural language processing methods. The objective is to provide a feature-based summary of a large number of health reviews of various doctors available online. Our experimental results indicate that the proposed techniques are very promising in performing these tasks. We sincerely believe that this problem will become increasingly important as more people are expressed their opinions on the Web. Summarizing of all will be useful to health consumers and also crucial to health service providers.

In future, we can extend this work to provide the aggregated summary of the information provided with large number of reviews available for given, respective provider. We plan to further improve and refine our techniques, and to deal with the outstanding problems such as the intensity of opinions, opinion changes over a period and investigating opinions expressed with adverbs, verbs and nouns.

REFERENCES

[1]. Kim, Hyun Duk, et al. "Comprehensive review of opinion summarization." (2011).

[2]. Minqing Hu and Bing Liu, Mining and Summarizing Customer Reviews.

[3]. Dave, K., Lawrence, S., and Pennock, D., 2003. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. *WWW'03*.

[4]. Morinaga, S., Ya Yamanishi, K., Tateishi, K, andFukushima, T. 2002. Mining Product Reputations on the Web.*KDD'02*. [18].

[5]. Cardie, C., Wiebe, J., Wilson, T. and Litman, D. 2003.Combining Low-Level and Summary Representations of Opinions for Multi-Perspective Question Answering. *2003 AAAI Spring Symposium on New Directions in Question Answering.*

[6]. Karlgren, J. and Cutting, D. 1994. Recognizing Text Genres with Simple Metrics using Discriminant Analysis.*COLING'94*.

[7]. Kessler, B., Nunberg, G., and Schutze, H. 1997. Automatic Detection of Text Genre. In *Proc. of 35th ACL/8th EACL*.

[8]. Finn, A., Kushmerick, N., and Smyth, B. 2002. Genre Classification and Domain Transfer for Information Filtering. In *Proc. of European Colloquium on Information Retrieval Research*, pages 353-362.

[9]. Goldstein, J., Kantrowitz, M., Mittal, V., and Carbonell, J.1999. Summarizing Text Documents: Sentence Selection and Evaluation Metrics. *SIGIR'99*.

[10]. Salton, G. Singhal, A. Buckley, C. and Mitra, M. 1996.Automatic Text Decomposition using Text Segments and Text Themes*. ACM Conference on Hypertext.*

[11]. Tait, J. 1983. *Automatic Summarizing of English Texts.* Ph.D. Dissertation, University of Cambridge

[12]. Hatzivassiloglou, V. and Wiebe, 2000. J. Effects of Adjective Orientation and Gradability on Sentence Subjectivity. *COLING'00*.

[13]. S. Das and M. Chen, "Yahoo! for Amazon: Extracting market sentiment from stock message boards," in *Proceedings of the Asia Pacific Finance Association Annual Conference (APFA)*, 2001.

[14]. Tong, R., 2001. An Operational System for Detecting and Tracking Opinions in on-line discussion. *SIGIR 2001 Workshop on Operational Text Classification*.

[15]. Hearst, M, 1992. Direction-based Text Interpretation as an Information Access Refinement. In Paul Jacobs, editor,*Text-Based Intelligent Systems*. Lawrence Erlbaum Associates.

[16]. Sack, W., 1994. On the Computation of Point of View.*AAAI'94*, Student abstract.

[17]. Turney, P. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *ACL'02*.

[18]. Pang, B., Lee, L., and Vaithyanathan, S., 2002. Thumbs up? Sentiment Classification Using Machine Learning Techniques. In *Proc. of EMNLP 2002*

[19]. Jacquemin, C., and Bourigault, D. 2001. Term extraction and automatic indexing. In R. Mitkov, editor, *Handbook of Computational Linguistics*. Oxford University Press.

[20]. Justeson, J. S., and Katz, S.M. 1995. Technical Terminology: some linguistic properties and an algorithm for identification in text. *Natural Language engineering*1(1):9-27.

[21]. Daille, B. 1996. Study and Implementation of Combined Techniques for Automatic Extraction of Terminology. *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*. MIT Press, Cambridge

[22]. Church, K.W. and Hanks, P. 1990. Word Association Norms, Mutual Information and Lexicography *Computational Linguistics*, 16(1):22-29

[23]. DeJong, G. 1982. An Overview of the FRUMP System*. Strategies for Natural Language Parsing*. 149-176

[24]. Tait, J. 1983. *Automatic Summarizing of English Texts.* Ph.D. Dissertation, University of Cambridge

[25]. Sparck J. 1993a. Discourse Modeling for Automatic Text Summarizing. *Technical Report 290,* University of Cambridge Computer Laboratory

[26]. Sparck J. 1993b.What might be in a summary? *Information Retrieval* 93:9-26.

[27]. Paice, C. D. 1990. Constructing Literature Abstracts by Computer: Techniques and Prospects. *Information Processing and Management* 26:171-186.

[28]. Kupiec, J., Pedersen, J., and Chen, F. 1995. A Trainable Document summarizer. *SIGIR'199*

[29]. Salton, G. Singhal, A. Buckley, C. and Mitra, M. 1996.Automatic Text Decomposition using Text Segments and Text Themes*. ACM Conference on Hypertext*

[30]. Mani, I., and Bloedorn, E., 1997. Multi-document Summarization by Graph Search and Matching. *AAAI'97*.

[31]. Boguraev, B., and Kennedy, C. 1997. Salience-Based Content characterization of Text Documents. In *Proc. Of the ACL'97/EACL'97 Workshop on Intelligent Scalable Text Summarization.*

[32]. P. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," in *Proceedings of the Association for Computational Linguistics (ACL)*, pp. 417–424, 2002.

[33]. B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 79–86,2002.

[34]. Wiebe, J. 2000. Learning Subjective Adjectives from Corpora. *AAAI'00*.

[35]. Wiebe, J., Bruce, R., and O'Hara, T. 1999. Development and Use of a Gold Standard Data Set for Subjectivity Classifications. In *Proc. of ACL'99*.

[36]. Bruce, R., and Wiebe, J. 2000. Recognizing Subjectivity: A Case Study of Manual Tagging. *Natural Language Engineering*.

[37]. Miller, G., Beckwith, R, Fellbaum, C., Gross, D., and Miller, K. 1990. Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography (special issue)*, 3(4):235-312.

[38]. Hatzivassiloglou, V. and Mckeown, K., 1997. Predicting the Semantic Orientation of Adjectives. In *Proc. of 35th ACL/8th EACL*.[41]. Fellbaum, C. 1998. *WordNet: an Electronic Lexical Database*, MIT Press.

[39]. NLProcessor–*Text AnalysisToolkit*.2000. http://www.infogistics.com/textanalysis.html

[40]. Huettner, A. and Subasic, P., 2000. Fuzzy Typing for Document management. In *ACL'00 Companion Volume:Tutorial Abstracts and Demonstration Notes*.

[41]. Manning, C. and Schütze, H. 1999. *Foundations of Statistical Natural Language Processing*, MIT Press.Cambridge, MA: May 1999.

[42]. Hu, M., and Liu, B. 2004. Mining Opinion Features in Customer Reviews. To appear in *AAAI'04*, 2004.

[43]. Liu, B., Hsu, W., Ma, Y. 1998. Integrating Classification and Association Rule Mining. *KDD'98*, 1998.

[44]. Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu, and Emery Jou Movie Rating and Review Summarization in Mobile Environment IEEE transactions on systems, man, and cybernetics—part c: applications and reviews, vol. 42, no. 3, may 2012 397

[45]. Radev, D. R., Hovy, E., and McKeown, K. (2002). Introduction to the special issue on summarization. Computational Linguistics., 28(4):399{408. [1, 2]

[46]. BALAHUR, A. AND MONTOYO, A. 2008. Multilingual feature-driven opinion extraction and summarization from customer reviews. In *NLDB '08: Proceedings of the 13th international conference on Natural Language and Information Systems*. Springer-Verlag, Berlin,Heidelberg, 345–346.

[47]. HU, M. AND LIU, B. 2006. Opinion extraction and summarization on the web. In *AAAI'06: proceedings of the 21st national conference on Artificial intelligence*. AAAI Press, 1621–1624.

[48]. LU, Y., ZHAI, C., AND SUNDARESAN, N. 2009. Rated aspect summarization of short comments. In *WWW '09:Proceedings of the 18th international conference on World wide web*. ACM, New York, NY, USA, 131–140.

[49]. KU, L.-W., LIANG, Y.-T., AND CHEN, H.-H. 2006. Opinion extraction, summarization and tracking in news and blog corpora. In *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*.100–107.

[50]. MEI, Q., LING, X., WONDRA, M., SU, H., AND ZHAI, C. 2007. Topic sentiment mixture: modeling facets and opinions in weblogs. In *WWW '07:* Proceedings *of the 16th international conference on World Wide Web*.ACM, New York, NY, USA, 171–180.

[51]. Radev, D. R., Hovy, E., and McKeown, K. (2002). Introduction to the special issue on summarization. Computational Linguistics. 28(4):399{408. [1, 2]