REF:JKU/2/83/022

Jomo Kenyatta University of Agriculture and Technology (JKUAT)

School of Computing and Information Technology (SCIT)

Department of Computing

Literature Review

Title: An Aspect-based Sentiment Analysis technique to predict product performance in promotional campaigns.

Student Name:KIRAGU PHILLIS WACERA Reg. No:CS282-0763/2011

Submission Date:23/03/2015

Supervisor 1:SYLVESTER KIPTOO

Supervisor 2:DR.AGNES MINDILA

Period: March 2015

Abstract

Textual information in the world can be broadly be categorized into two main types: facts and opinions. Facts are objective expressions about entities, events and their properties. Opinions are usually subjective expressions that describe people's sentiments, appraisals or feelings toward entities, events and their properties. Much of the existing research on textual information processing has been focused on mining and retrieval of factual information, e.g., information retrieval, Web search, text classification, text clustering and many other text mining and natural language processing tasks. Little work had been done on the processing of opinions until only recently. Yet, opinions are so important that whenever we need to make a decision we want to hear others opinions. This is not only true for individuals but also true for organizations.

This document discuss sentiment analysis, the field that is considered with opinion mining from user generated content. It describes the challenges, techniques and application of sentiment analysis. It will concentrate on aspect based analysis which is a sentiment analysis technique that analyses individual phrases in sentences in order to determine the orientation, that is, positive neutral and negative. It will highlight the research that has been done by others in this field, the algorithms used for the analysis their advantages and challenges faced by each of them.

Contents

1 Chapter 2:LITERATURE REVIEW

1.1 Introduction

Sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. It is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes,and emotions towards entities such as products, services, organizations,individuals, issues, events, topics, and their attributes.(Liu 2012)

Sentimental Analysis is all about to get the real voice of people towards specific product, services, organization, movies, news, events, issues and their attributes(liu, 2012). Sentiment Analysis includes branches of computer science like Natural Language Processing, Machine Learning, Text Mining and Information Theory and Coding. By using approaches, methods, techniques and models of defined branches, we can categorized our data which is unstructured data may be in form of news articles, blogs, tweets, movie reviews, product reviews etc. into positive, negative or neutral sentiment according to the sentiment is expressed in them. Sentiment analysis is done on four levels (liu, 2012): Document Level, Sentence Level, Comparative level, Entity or Aspect Level.

Document Level Sentiment analysis is performed for the whole document and then decide whether the document express positive or negative sentiment. (liu, 2012)

Comparative statements are like the entity or aspect level sentiment analysis but deal with techniques of comparative sentiment analysis.

Sentence level sentiment analysis is related to find sentiment form sentences whether each sentence expressed a positive, negative or neutral sentiment. Sentence level sentiment analysis is closely related to subjectivity classification.

Entity or Aspect Level sentiment analysis performs finer-grained analysis. The goal of entity or aspect level sentiment analysis is to find sentiment on entities and/or aspect of those entities. For example consider a statement "My phone has good picture quality but it has low phone memory storage." so sentiment on the phone's camera and display quality is positive but the sentiment on its phone memory storage is negative. We can generate summary of opinions about entities.

Following the above techniques, Researchers in sentiment analysis have focused mainly on two problems: detecting whether the text is subjective or objective, and determining whether the subjective text is positive or negative. These techniques relied on two main approaches: unsupervised sentiment orientation calculation, and supervised and unsupervised classifications based on machine learning. (meng, 2012)

1.2 Approaches of sentiment analysis

1.2.1 Machine Learning Techniques

Machine learning techniques are most useful techniques for the sentiment analysis for categorized document or sentences into positive, negative or neutral categories. Machine learning techniques classified into two basic techniques as defined below. (Indurkhya et al,2010, Feldman, 2007, han et al, 2006)

Supervised Machine Learning Techniques: Supervised machine learning techniques are used for classified document or sentences into finite set of class i.e into positive, negative and neutral. Training data set is available for all kind of classes. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances. This requires the learning algorithm to generalize from the training data to unseen situations in a ”reasonable” way.(wikipedia) The supervised machine learning classifications mainly reside in document-level classification and sentence-level classification.(weng,2012)

Unsupervised Machine Learning Techniques: Unsupervised machine learning techniques don't use training data set for classification. Clustering algorithms like K-means clustering, Hierarchical clustering used to classify data into categories. Semantic Orientation also provides to generate accurate result for classification. Neural network can be also used for defining threshold values to the words and classify them based on the defined values. Point wise mutual information (PMI) is also one of the unsupervised classification methods for sentiment analysis.

1.3 Techniques involved in Sentiment Analysis

There are four main techniques:(Liu 2012)

1.3.1 Document-Level Sentiment Analysis

The task at this level is to classify whether a whole opinion document expresses a positive or negative sentiment (Turney, 2002). For example, given a product review, the system determines whether the review expresses an overall positive or negative opinion about the product. This task is commonly known as document-level sentiment classification. This level of analysis assumes that each document expresses opinions on a single entity (e.g., a single product). Thus, it is not applicable to documents which evaluate or compare multiple entities.

In a document-level classification, a document can be classified based on the frequency of different words in the Bag-of-Word. Before supervised machine learning classification methods can be applied, document-level analysis needs the pre-labeled training data such as Thumb-up or Thumb-down labels (Pang et al., 2002; Turney, 2002). The problem with document-level analysis is that we cannot get more detailed information such as positive/negative sentiments regarding certain product features from the reviews. For example, a product such as a car consists several product features like engines, tires, and batteries. Most of the time, one person can express his/her opinion on more than one product features in the same review or even in the same sentence. For example, "I like the size, but its battery life is short." So, it is meaningful to conduct sentiment analysis at a more detailed level.

Sentiment classification at the document level provides an overall opinion on an entity, topic or event. It has been studied by a large number of researchers. However, this level of classification has some shortcomings for applications:(liu 2012)

Shortcomings of document level sentiment analysis

In many applications, the user needs to know additional details, e.g., what aspects of entities are liked and disliked by consumers. In typical opinion documents, such details are provided, but document sentiment classification does not extract them for the user.

Document sentiment classification is not easily applicable to non-reviews such as forum discussions, blogs, and news articles, because many such postings can evaluate multiple entities and compare them. In many cases, it is hard to determine whether a posting actually evaluates the entities that the user is interested in, and whether the posting expresses any opinion at all, let alone to determine the sentiment about them. Document-level sentiment classification does not perform such finegrained tasks, which require in-depth natural language processing. In fact, online reviews do not need sentiment classification because almost all reviews already have user-assigned star ratings. In practice, it is the forum discussions and blogs that need sentiment classification to determine people's opinions about different entities (e.g., products and services) and topics.

1.3.2 Sentence-Level Sentiment Analysis

A single document may contain multiple opinions even about the same entities. When there is a need to have a more fine-grained view of the different opinions expressed in the document about the entities we must analyse the sentences within the document.

The task at this level goes to the sentences and determines whether each sentence expressed a positive, negative, or neutral opinion. Neutral usually means no opinion. This level of analysis is closely related to subjectivity classification which distinguishes sentences (called objective sentences) that express factual information from sentences (called subjective sentences) that express subjective views and opinions. However, we should note that subjectivity is not equivalent to sentiment as many objective sentences can imply opinions, e.g., "I bought this camera last week and the shutter is broken" Researchers have also analyzed clauses but the clause level is still not enough, e.g., "Lenovo is doing very well in this lousy economy."(liu,2012)

Shortcoming of Sentence-Level Sentiment Analysis

Sentence level subjectivity classification and sentiment classification goes further than document level sentiment classification as it moves closer to opinion targets and sentiments on the targets. It can be regarded as an intermediate step in the overall sentiment analysis task. However, it still has several shortcomings for many real-life applications:(liu, 2012)

In most applications, the user needs to know additional details, i.e., what entities or aspects of entities are liked and disliked. As the document level, the sentence level analysis still does not do that.

Although one may say that if we know the opinion targets (e.g., entities and aspects, or topics), we can assign the sentiment orientation of a sentence to the targets in the sentence. However, this is insufficient:

(1) Many complex sentences have different sentiments on different targets, e.g., "Trying out Chrome because Firefox keeps crashing" and "Lenovo is doing very well in this lousy economy." In this latter sentence, even the clause level classification is insufficient. We need to go to the opinion target or the aspect level.

(2) Although a sentence may have an overall positive or negative tone, some of its components may express opposite opinions. For example, some researchers regard the follow sentence as positive (Neviarouskaya et al., 2010; Zhou et al., 2011):

"Despite the high unemployment rate, the economy is doing well." It is true that the overall tone of this sentence is positive or the author is trying to emphasize the positive side, but it does contain a negative sentiment on the unemployment rate, which we must not ignore. If we go to the aspect-level sentiment analysis, the problem is solved. That is, the sentence is positive about the overall economy but negative about the unemployment rate.

(3) Sentence level sentiment classification cannot deal with opinions in comparative sentences, e.g., "Coke tastes better than Pepsi." In this case, we need different methods to extract and to analyze comparative opinions as they have quite different meanings from regular opinions. Although this sentence clearly expresses an opinion, we cannot simply classify the sentence as being positive, negative or neutral.

1.3.3 Comparative Sentiment Analysis

In many cases users do not provide a direct opinion about one product but instead provide comparable opinions such as in these sentences taken from the user forums of Edmonds.com that compares different car models: "300 C touring looks so much better than the Magnum," "I drove the Honda Civic, it does not handle better than the TSX, not even close." The goal of the sentiment analysis system in this case is to identify the sentences that contain comparative opinions, and to extract the preferred entity(ies) in each opinion. One of the pioneering papers on comparative sentiment analysis is Jindal and Liu(2006). This paper found that using a relatively small number of words we can cover 98

Comparative adjectives adverbs such as: 'more,' 'less,' and words ending with -er (for example, 'lighter').

Superlative adjectives and adverbs such as: 'most,' 'least,' and words ending with -est (for example, 'finest').

Additional phrases such as 'favor,' 'exceed,' 'outperform,' 'prefer,' 'than,' 'superior,' 'inferior,' 'number one,' 'up against.' Since these words lead to a very high recall, but low precision, a naive Bayes classifier can be used to filter out sentences that do not contain comparative opinions. The classifier uses sequential patterns as features. The sequential patterns are discovered by the class sequential rule (CSR) mining algorithm.

Although there have been some existing works, comparative sentences have not been studied as extensively as many other topics of sentiment analysis. Further research is still needed. One of the difficult problems is how to identify many types of non-standard or implicit comparative sentences, e.g., “I am very happy that my iPhone is nothing like my old ugly Droid.” Without identifying them, further sentiment analysis is hard to perform.

1.3.4 Aspect-Based Sentiment Analysis

The two previous approaches work well when either the whole document or each individual sentence refers to a single entity. However, in many cases people talk about entities that have many aspects (attributes) and they have a different opinion about each of the aspects. This often happens in reviews about products or in discussion forums dedicated to specific product categories (such as cars, cameras, smartphones, and even pharmaceutical drugs). As an example here is a review of Kindle Fire taken illustrated by Jindal and Liu: ”As a long-time Kindle fan I was eager to get my hands on a Fire. There are some great aspects; the device is quick and for the most part dead-simple to use. The screen is fantastic with good brightness and excellent color, and a very wide viewing angle. But there are some downsides too; the small bezel size makes holding it without inadvertent page-turns difficult, the lack of buttons makes controls harder, the accessible storage memory is limited to just 5 GB.”

Classifying this review as either positive or negative toward the Kindle would totally miss the valuable information encapsulated in it. The author provides feedback about many aspects of the Kindle (like speed, ease of use, screen quality, bezel size, buttons, and storage memory size). Some of these aspects are reviewed positively while some of the others get a negative sentiment. Aspect-based sentiment analysis (also called feature-based sentiment analysis) is the research problem that focuses on the recognition of all sentiment expressions within a given document and the aspects to which they refer.

This study will concentrate on the aspect based sentiment analysis. it has two core tasks that have been extensively researched by many researchers as quoted by Bing Liu in his paper.(liu, 2012). these core tasks are

- Aspect extraction: This task extracts aspects that have been evaluated.For example, in the sentence, ”The image quality of this Camera is amazing,” the aspect is ”image quality” of the entity represented by ”this camera.” Note that ”this camera”does not indicate the aspect GENERAL here because the evaluation is not about the camera as a whole, but only about its voice quality. However, the sentence ”I love this camera” evaluates the camera as a whole, i.e., the GENERAL aspect of the entity represented by ”this camera.” Bear in mind whenever we talk about an aspect, we must know which entity it belongs to.

- Aspect sentiment classification: This task determines whether the opinions on different aspects are positive, negative, or neutral. In the first example above, the opinion on the ”voice quality” aspect is positive. In the second, the opinion on the aspect GENERAL is also positive.

Aspect extraction

In the context of sentiment analysis, some specific characteristics of the problem can facilitate the extraction. The key characteristic is that an opinion always has a target. The target is often the aspect or topic to be extracted from a sentence.(Pontiki et al, 2014) Thus, it is important to recognize each opinion expression and its target from a sentence. However, we should also note that some opinion expressions can play two roles, i.e., indicating a positive or negative sentiment and implying an (implicit) aspect (target). For example, in ”this car is expensive,” ”expensive” is a sentiment word and also indicates the aspect price. Here, we will focus on explicit aspect extraction. There are four main approaches: (liu, 2012)

- Extraction based on frequent nouns and noun phrases.
- Extraction by exploiting opinion and target relations.
- Extraction using supervised learning.
- Extraction using topic modeling.

Since existing research on aspect extraction (more precisely, aspect expression extraction) is mainly carried out in online reviews, we also use the review context to describe these techniques, but there is nothing to prevent them being used on other forms of social media text. There are two common review formats on the Web.

- Format 1 Pros, Cons, and the detailed review: The reviewer first describes some brief pros and cons separately and then writes a detailed/full review.
- Format 2 Free format The reviewer writes freely, i.e., no brief pros and cons. Extracting aspects from Pros and Cons in reviews of Format 1 (not the detailed review, which is the same as that in Format 2) is a special case of extracting aspects from the full review and also relatively easy. In (Liu, Hu and Cheng, 2005), a specific method based on a sequential learning method was proposed to extract aspects from Pros and Cons, which also exploited a

key characteristic of Pros and Cons, i.e., they are usually very brief, consisting of short phrases or sentence segments. Each segment typically contains only one aspect. Sentence segments can be separated by commas, periods, semi-colons, hyphens, and, but, etc. This observation helps the extraction algorithm to perform more accurately.

This study will concentrate on extraction based on frequent nouns and noun phrases.

Finding Frequent Nouns and Noun Phrases

This method finds explicit aspect expressions that are nouns and noun phrases from a large number of reviews in a given domain. Hu and Liu (2004) used a data mining algorithm. Nouns and noun phrases (or groups) were identified by a part-of-speech (POS) tagger. Their occurrence frequencies are counted, and only the frequent ones are kept. A frequency threshold can be decided experimentally. The reason that this approach works is that when people comment on different aspects of an entity, the vocabulary that they use usually converges. Thus, those nouns that are frequently talked about are usually genuine and important aspects. Irrelevant contents in reviews are often diverse, i.e., they are quite different in different reviews. Hence, those infrequent nouns are likely to be non-aspects or less important aspects. Although this method is very simple, it is actually quite effective. Some commercial companies are using this method with several improvements. The precision of this algorithm was improved in (Popescu and Etzioni, 2005). Their algorithm tried to remove those noun phrases that may not be aspects of entities. It evaluated each discovered noun phrase by computing a pointwise mutual information (PMI) score between the phrase and some meronymy discriminators associated with the entity class, e.g., a camera class. The meronymy discriminators for the camera class are, “of camera,” “camera has,” “camera comes with,” etc., which were used to find components or parts of cameras by searching the Web. The PMI measure is as follows:

$$PMI(a, d) = hits(a \wedge d) \mid hits(a)hits(d) \tag{1}$$

where a is a candidate aspect identified using the frequency approach and d is a discriminator. Web search was used to find the number of hits of individual terms and also their co-occurrences. The idea of this approach is clear. If the PMI value of a candidate aspect is too low, it may not be a component of the product because a and$d$do not co-occur frequently. The algorithm also distinguishes components/parts from attributes using WordNet’s is-a hierarchy (which enumerates different kinds of properties) and morphological cues (e.g., “-iness,” “-ity” suffixes).

Blair-Goldensohn et al. (2008) refined the frequent noun and noun phrase approach by considering mainly those noun phrases that are in sentiment bearing sentences or in some syntactic patterns which indicate sentiments. Several filters were applied to remove unlikely aspects, e.g., dropping aspects which do not have sufficient mentions along-side known sentiment words. They also collapsed aspects at the word stem level, and ranked the discovered aspects by a manually tuned weighted sum of their frequency in sentiment-bearing sentences and the type of sentiment phrases/patterns, with appearances in phrases carrying a greater weight. U

A frequency-based approach was also taken in (Ku, Liang and Chen, 2006). The authors called the so discovered terms the major topics. Their method also made use of the TF-IDF scheme considering terms at the document level and at the paragraph level. Moghaddam and Ester (2010) augmented the frequency-based approach with an additional pattern-based filter to remove some non-aspect terms. Their work also predicted aspect ratings. Scaffidi et al. (2007) compared the frequency of extracted frequent nouns and noun phrases in a review corpus with their occurrence rates in a generic English corpus to identify true aspects.

Zhu et al. (2009) proposed a method based on the Cvalue measure from (Frantzi, Ananiadou and Mima, 2000) for extracting multi-word aspects. The Cvalue method is also based on frequency, but it considers the frequency of multi-word term t, the length of t, and also other terms that contain t. However, Cvalue only helped find a set of candidates, which is then refined using a bootstrapping technique with a set of given seed aspects. The idea of refinement is based on each candidate’s co-occurrence with the seeds. Long, Zhang and Zhu (2010) extracted aspects (nouns) based on frequency and information distance. Their method first finds the core aspect words using the frequency-based method. It then uses the information distance in (Cilibrasi and Vitanyi, 2007) to find other related words to an aspect, e.g., for aspect price, it may find “dollars”. All these words are then used to select reviews which discuss a particular aspect most.

Sentiment Classification

There are two main approaches in determining the orientation of a sentiment, i.e., the supervised learning approach and the lexicon-based approach.

However, the key issue is how to determine the scope of each sentiment expression, i.e., whether it covers the aspect of interest in the sentence. The current main approach is to use parsing to determine the dependency and the other relevant information.

Supervised learning is dependent on the training data. A model or classifier trained from labeled data in one domain often performs poorly in another domain. Although domain adaptation (or transfer learning) has been studied by researchers, the technology is still far from mature, and the current methods are also mainly used for document level sentiment classification as documents are long and contain more features for classification than individual sentences or clauses. Thus, supervised learning has difficulty to scale up to a large number of application domains. The lexicon-based approach can avoid some of the issues (Ding, Liu and Yu, 2008; Hu and Liu, 2004), and has been shown to perform quite well in a large number of domains. Such methods are typically unsupervised. They use a sentiment lexicon (which contains a list of sentiment words, phrases, and idioms), composite expressions, rules of opinions, and (possibly) the sentence parse tree to determine the sentiment orientation on each aspect in a sentence. They also consider sentiment shifters, but-clause and many other constructs which may affect sentiments. Discussed below is the lexicon-based method (Taboada et al,2014) and it has four steps. Here, we assume that entities and aspects are known:

1. Mark sentiment words and phrases: For each sentence that contains one or more aspects, this step marks all sentiment words and phrases in the sentence. Each positive word is assigned the sentiment score of +1 and each negative word is assigned the sentiment score of 1. For example, the sentence, "The voice quality of this phone is not good, but the battery life is long." After this step, the sentence becomes "The voice quality of this phone is not good [+1], but the battery life is long" because "good" is a positive sentiment word (the aspects in the sentence are italicized). Note that "long" here is not a sentiment word as it does not indicate a positive or negative sentiment by itself in general, but we can infer its sentiment in this context shortly.

2. Apply sentiment shifters: Sentiment shifters (also called valence shifters) are words and phrases that can change sentiment orientations. There are several types of such shifters. Negation words like not, never, none, nobody, nowhere, neither, and cannot are the most common type. This step turns our sentence into "The voice quality of this phone is not good [-1], but the battery life is long" due to the negation word "not."

3. Handle but-clauses: Words or phrases that indicate contrary need special handling because they often change sentiment orientations too. Sentiment Analysis and Opinion Mining 61) The most commonly used contrary word in

English is “but”. A sentence containing a contrary word or phrase is handled by applying the following rule: the sentiment orientations before the contrary word (e.g., but) and after the contrary word are opposite to each other if the opinion on one side cannot be determined. The if-condition in the rule is used because contrary words and phrases do not always indicate an opinion change, e.g., “Car-x is great, but Car-y is better.” After this step, the above sentence is turned into “The voice quality of this phone is not good [-1], but the battery life is long [+1]” due to “but” ([+1] is added at the end of the but-clause). Notice here, we can infer that “long” is positive for “battery life”. Apart from but, phrases such as “with the exception of,” “except that,” and “except for” also have the meaning of contrary and are handled in the same way. As in the case of negation, not every but means contrary, e.g., “not only ... but also.” Such non-but phrases containing “but” also need to be identified beforehand.

4. Aggregate opinions: This step applies an opinion aggregation function to the resulting sentiment scores to determine the final orientation of the sentiment on each aspect in the sentence. Let the sentence be s, which contains a set of aspects $[a_1, \ldots, a_m\ ]$ and a set of sentiment words or phrases $[sw_1, \ldots, sw_n]$ with their sentiment scores obtained from steps 1-3. The sentiment orientation for each aspect a_i in s is determined by the following aggregation function:

$$score(a_i, s) = \sum_{ow_j \in s} sw_j.so \mid dist(sw_j, a_i) \tag{2}$$

where sw_j is an sentiment word/phrase in s, $dist(sw_j, a_i)$ is the distance between aspect a_i and sentiment word sw_j in s. $sw_j.so$ is the sentiment score of sw_i. The multiplicative inverse is used to give lower weights to sentiment words that are far away from aspect a_i. If the final score is positive, then the opinion on aspect a_i in s is positive. If the final score is negative, then the sentiment on the aspect is negative. It is neutral otherwise.(liu, 2012)

This simple algorithm performs quite well in many cases. It is able to handle the sentence “Apple is doing very well in this bad economy” with no problem. Note that there are many other opinion aggregation methods. For example, (Hu and Liu, 2004) simply summed up the sentiment scores of all sentiment words in a sentence or sentence segment. Kim and Hovy used multiplication of sentiment scores of words. Similar methods were also employed by other researchers wan and zhu.(liu, 2012)

Along with aspect sentiment classification research, researchers also studied the aspect sentiment rating prediction problem which has mostly been done together with aspect extraction in the context of topic modeling.

2 Chapter 3:RESEARCH METHODOLOGY

2.1 Introduction

This study extracted user sentiments from online chat users. The data was the annotated to extract related chats. The study was focused on internet users who review products. In this case several reviews were obtained from the website productsreview.com, where users of different products review products.

The participants in this study are people who have used the product and done online reviews on the same. The reviewers all have similar phone type: Samsung Galaxy S

2.2 Data Collection Techniques

The data collected in the study will be in text form. It will contain comments, arguments and sentiments expressed by people about a product.

The topic was: HOW DO YOU FEEL ABOUT THE SAMSUNG GALAXY S. Below are some of the comments:

Antony

”Best phone around... love it to bits!!!!! I Waiting for a while before upgrading to this...just wow does everything. The screen is so great... and the pictures it takes and the 4k recording is crazy good. I do recommend this to everyone who is sitting on the fence about this phone... it is worth it!”

Nina Rose

”Being told it was the best phone on the market I obliged and trusted their opinion. I will never purchase another samsung phone again!!! Grossly over rated!! IPHONE BY FAR THE BEST!!! Not responsive!!! Too large....waste of my time and hard earned money!!!! Key pad so badly set out!!! Everything is wrong with this phone! Hate it with a passion!!!!”

Renee

”It is absolutely the worst phone I have ever owned. So many problems that I don’t know where to start. It continuously overheats whilst charging or talking. It freezes during calls and you’re unable to end a call, retrieve it or do anything else with the phone except restart it. The touch screen decides to stop working intermittently, so again you’re unable to make the phone function at all. The volume function only works sporadically, ie it's supposed to be on silent but still makes noise. I sent the phone back to be repaired but was told there's nothing wrong with it. It's quite astounding to think that they're supposed to be such poor in quality. To top it all off, they're not even user friendly...far more difficult to navigate than they should be.But the phone is a good size and easy to use .I have had a galaxy ace and this is a little quicker .I haven't gone to all of the icons and found out what they do but will get there eventually . I like the weather app. and have it on the home page and check it often before planning the day"

2.3 Data Analysis

This study will use unsupervised learning to classify the opinions into their orientation. That is whether positive negative or neutral. Discussed below is the lexicon-based method and it has four steps.

2.3.1 Taking the first review:

1. Mark sentiment words and phrases: For each sentence that contains one or more aspects, this step marks all sentiment words and phrases in the sentence. Each positive word is assigned the sentiment score of +1 and each negative word is assigned the sentiment score of -1.

"Best phone around... love it to bits!!!!! I Waiting for a while before upgrading to this...just wow does everything. The screen is so great... and the pictures it takes and the 4k recording is crazy good. I do recommend this to everyone who is sitting on the fence about this phone... it is worth it! "

sentence	Sentiment phrase	score
1	best	+1
1	love	+1
3	wow	+1
4	great	+1
5	crazy	-1
5	good	+1
6	recommend	+1
6	worth	+1

2. Apply sentiment shifters: Sentiment shifters (also called valence shifters in (Polanyi and Zaenen, 2004)) are words and phrases that can change sentiment orientations. There are several types of such shifters. Negation words like not,

never, none, nobody, nowhere, neither, and cannot are the most common type. This changes the sentiment crazy good into one +1 instead of two separate sentiments.

sentence	Sentiment phrase	score
1	best	+1
1	love	+1
3	wow	+1
4	great	+1
5	crazy good	+1
6	recommend	+1
6	worth	+1

3. Handle but-clauses: We skip this step as there are no but clauses in this sentiment

4. Aggregate opinions: We get the aggregated opinion by using addition as used by hu and liu ,2004 The review is positive with +7

2.3.2 Taking the third review:

1. Mark sentiment words and phrases: For each sentence that contains one or more aspects, this step marks all sentiment words and phrases in the sentence. Each positive word is assigned the sentiment score of +1 and each negative word is assigned the sentiment score of -1.

"It is absolutely the worst phone I have ever owned. So many problems that I don't know where to start. It continuously overheats whilst charging or talking. It freezes during calls and you're unable to end a call, retrieve it or do anything else with the phone except restart it. The touch screen decides to stop working intermittently, so again you're unable to make the phone function at all. The volume function only works sporadically, ie it's supposed to be on silent but still makes noise. I sent the phone back to be repaired but was told there's nothing wrong with it. It's quite astounding to think that they're supposed to be such poor in quality. To top it all off, they're not even user friendly...far more difficult to navigate than they should be.But the phone is a good size .I have had a galaxy ace and this is a little quicker . I like the weather app. and have it on the home page and check it often before planning the day"

sentence	Sentiment phrase	score
1	worst	-1
1	problems	-1
3	overheats	-1
4	freezes	-1
5	sporadically	-1
5	poor	-1
5	user friendly	+1
6	difficult	-1
6	good	+1
7	quicker	+1
8	like	+1

2. Apply sentiment shifters: Sentiment shifters (also called valence shifters in (Polanyi and Zaenen, 2004)) are words and phrases that can change sentiment orientations. There are several types of such shifters. Negation words like not, never, none, nobody, nowhere, neither, and cannot are the most common type. This applies to the sentiment not user friendly

sentence	Sentiment phrase	score
1	worst	-1
1	problems	-1
3	overheats	-1
4	freezes	-1
5	sporadically	-1
5	poor	-1
5	user friendly	-1
6	difficult	-1
6	good	+1
7	quicker	+1
8	like	+1

3. Handle but-clauses: The but recognizes positive sentiments.

4. Aggregate opinions: We get the aggregated opinion by using addition as used by hu and liu ,2004 -8+3= +5

2.4 Conclusion

It is clear that people who have an experience with a new or existing product get to express their opinion of it in an unstructured way. This expressions carry sentiments with them that can be analyzed and then used to project the performance of a product for a target market in a promotional campaign. Thus, this study will provide a platform through which such users can obtain the sentiments, analyze and classify them to determine their orientation that is, whether it's positive, negative or neutral. using the algorithm showcased above it is possible to rate sentiments according to their orientation

References

Liu, b. *Sentiment analysis and opinion mining.*, Synthesis Lectures on Human Language Technologies. Morgan and Claypool Publishers, 2012.

Meng, Yanyan. *"SENTIMENT ANALYSIS: A Study On Product Features"*, Dissertations and Theses from the College of Business Administration. Paper 28, 2012.

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

Nitin Indurkhya, Fred J. Damerau. *Handbook of Natural Language Processing.* , Second Edition, CRC Press, 2010.

Taboada, M. , Brooke, J. Tofiloski ,M. and Voll, K. *Lexicon-Based Methods for Sentiment Analysis*, university of Toronto , 2014.

Turney, P. *Thumbs up or thumbs down?*,Semantic orientation applied to unsupervised classification of reviews. In Proceedings of the Association for Computational Linguistics, 2002.

Pontiki, M. , Galanis, D. , Pavlopoulos, J. and Manandhar,s. *Aspect Based Sentiment Analysis*,SemEval-2014 Task 4.

Narayanan, r., Liu, b. and Chaudhary, a. *Sentiment analysis of conditional sentences.*, In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (Singapore), association for Computational Linguistics, 2009.

Tsur, o., Davidov, d. and Rappoport, a. *A great catchy name: semi-supervised recognition of sarcastic sentences in online product reviews.*, In Fourth International AAAI Conference on Weblogs and Social Media, 2010.

Jindal, n. and Liu, b. *identifying comparative sentences in text documents.*, In Proceedings of ACM SIGIR Conf. on Research and Development in Information Retrieval,2014.

Popescu, a.-m. and Etzioni, o. *Extracting product features and opinions from reviews.*, In Proceedings of Conference on Empirical Methods in Natural Language Processing, 2005.

Ding, X., Liu, b. and Zhang, L. *Entity discovery and assignment for opinion mining applications.*, in Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining , 2009.

Ronen Feldman, James Sanger. *The Text Mining Handbook-Advance Approaches in Analyzing Unstructured Data* , Cambridge University Press,2007. Han et al,2006 Jiawei Han, Micheline Kamber, *Data Mining Concepts and Techniques*, Morgan Kaufmann Publications, 2006.

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

Neviarouskaya, Alena, Helmut Prendinger, and Mitsuru Ishizuka. *Recognition of affect, judgment, and appreciation in text.*, in Proceedings of the 23rd International Conference on Computational Linguistics (COLING-2010). 2010.

Zhou, Lanjun, Binyang Li, Wei Gao, Zhongyu Wei, and Kam-Fai Wong. *Unsupervised discovery of discourse relations for eliminating intrasentence polarity ambiguities.*, in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2011). 2011. Wikipedia article on supervised machine learning http://en.m.wikipedia.org/wiki/Supevised$_learning$.