International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 14 (2017) pp. 4528-4533

Expeditious Approach to Identify the Product Sentiment Based on MORQA System and Polarity Shift Mechanism

Vallikannu Ramanathan[1] and T.Meyyappan[2]

[1]*Research Scholar, Alagappa University, Karaikudi, Tamilnadu, India.*

[1]*Orcid: 0000-0002-5828-9631*

[2]*Professor, Alagappa University, Karaikudi, Tamilnadu, India.*

Abstract

Online reviews are often first impression for the consumers to purchase the product through online. When estimating the feasible purchase, we may have a query in mind e.g., 'Will this speaker is louder enough for the party hall?' or 'Should I buy iPhone 7S or iPhone 7S+?' To answer these queries we may have to surfing the online consumer review sites , either to read the large volume of consumer reviews which relevant to an item or post our product related queries to the people community via Question & Answer (Q/A) system.

In this paper, we combine both paradigms: consumer product review system and consumer Q/A system to get the best analysis about the product. We articulate this as a machine learning problem using our proposed new system called Mixture of Reviews and Questions & Answers (MORQA) to create a framework and to find the opinion about the online products. Our idea is to identify the opinion based on MORQA system and resolve the negation problem in case of negative sentences in the opinion context. Negation problem is the major issue which affects the result of the sentiment analysis. Sentiment analysis lexicon feature and antonym thesaurus are used to resolve this problem. We also provide a visualization for our result summarization.

Keywords: Text mining; Sentiment Analysis; Consumer online review; Question Answering; MORQA; opinion lexicon; negation; product features lexicon.

INTRODUCTION

Information can be classified into truths and opinions. Truth is an information that can written with evidence. Opinion is based on a user's view or belief, it might be supported by facts. Before World Wide Web (WWW) we were deficient to collect large volume of opinions about the product. Earlier people collect opinion about the product through friends and relatives from their word of mouth. After introducing WWW people can post their product reviews in trader's web sites and express their views about the product in internet environments, blogs and chats.

Online shopping has become essential part of people life. There are many websites offer online shopping facilities to the consumers. These websites provide platform on which consumers can share their opinions about the products. Amazon is the one of the most important online shopping website in the world. Consumers can register their review about the product in amazon. The amazon consumer review feedback dataset and consumer Q/A system used as input to predict the sentiment of customer about the product. To achieve this bag of words and bigrams are used.

Amazon reviews are not only about the product, also discuss about the features of the product and services given to the consumers.

In text mining, many methods and models are used to classify the text. Sentiment analysis as a part of text mining used to identify sentiment or opinion about any product. Sentiment analysis on product review has been used to improve business targets and customer retention.

RELATED WORK

Sentiment analysis or opinion mining has scored more attraction in recent years. Sentiment analysis plays a very crucial role in online shopping [5]. Sentiment analysis is a study of people's opinion, attitude and emotions about particular object or thing [5].

Sentiment analysis is a special text mining task used to determine and extract subjective attitude or sentiment of people from a given text [3][7]. Sentiment classification is a basic task in sentiment analysis to classify people's sentiment which is expressed in text format into different polarity classes [1].

The most popular text representation model in machine learning based sentiment classification is known as bag-of-words (BOW) model [2]. Although the BOW model is simple and has achieved great successes in topic-based text classification [2]. BOW model is a traditional model used to break the sentences into multi-set of words based on which standard machine learning algorithms are employed as classifiers.

Negations problem or polarity shift is one of the tedious problem to manage in sentiment analysis and classification. Polarity shift is a linguistic phenomenon which can shift or inverse the sentiment of the given text. For example we may have a positive review "I like this Sony C500 camera's optical zoom feature". Polarity shift of this sentence will become reverse review of original review by adding negation word "don't" in front of the opinion word "like", "I don't like this Sony C500 camera's optical zoom feature". Here we can see sentiment orientation is gets converted from positive to negative.

Opinion is the thought, feeling or judgment given by the consumers. Opinion mining is the study of people's opinion towards certain entities [1].

In this paper, we aim to categorize sentiment polarity of online site reviews, which is one of the important factor for the people to purchase an items through online shopping. Reviews, questions & answers of amazon ecommerce site is used as data for this research. Online shopping success rate depends on star rating, consumer review posting and question & answer system.

We retrieve review rating of the product from the consumers after they utilized the product. The rating allocated to the product by the consumer denoted by a number of stars on a scale 1 to 5. In general opinion can be classified into positive, negative or neutral. We propose this can be further segregated into five categories: extremely positive, positive, neutral, negative or extremely negative. These five categories of sentiment can be compared with five levels of star rating given by consumers in amazon such as very satisfied, satisfied, neutral, disappointed or very disappointed.

TABLE 1: Opinion Classification

Star Level	Meaning
☆	very disappointed
☆☆	Disappointed
☆☆☆	Neutral
☆☆☆☆	Satisfied
☆☆☆☆☆	very satisfied

SYSTEM DESIGN

Amazon star ratings do not express the exact feeling of the consumers. Generally large amount of reviews (80%) fall in the range of five to four stars and very few review ratings fall below four stars. Star ratings and reviews posted by consumers are not exactly express the same type of opinion. Some cases consumer review statements opinion are slightly different from the star rating opinion. Typically consumer reviews made of simple or complex sentences which prompt the mixed emotions of positive opinions as well as negative opinions about the product features. Our model intensified to find the opinion of the product based on consumer reviews and consumer Q/A system in addition to that of star rating.

Our proposed model uses bag-of-word feature to extract sentiment from consumer reviews. In addition to improve the accuracy adverb and adjective features are combine with bag of words. This combined features increase the accuracy of the sentiment result.

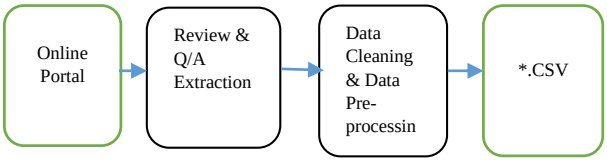

Figure 1: Online product review analysis

A. Data Pre-processing

Crawl the amazon reviews to extract the specific product related queries from the web portal. Pre-processing of data plays very crucial role in extracted product reviews. Data pre-processing is the first step which involves tokenization, data filtering, stop word removal and stemming.

B. Data Cleaning

Data cleaning to be done to remove stop words and special characters. The extracted data to be cleaned so that we can analysis the review efficiently. Cleaning of crawled data is done by the following way.

- Removing all special characters (such as '/', ':', '?', '&', '-')
- Repeated letters (e.g., DELL this model laptop is soooooooooo good.) must be cleaned so that any letter occurring more than one time is replaced with single occurrence of the letter. In the above e.g. sooooooo must be converted to 'so'.
- Stop words must be filtered from the review. 'the', 'is', 'at', 'of', 'on' and 'a' are the few examples of the stop words.

After cleaning, the crawled content must be stored in CSV file.

C. Opinion lexicon

Here we use external lexicons that endorse synonyms and lexicon based antonyms for sentiment words. Our method uses Hu&Liu's opinion lexicon for identifying sentiment words and WordNet dictionary to find antonyms. WordNet dictionary maintains the set of lexical database for English words. The proposed method uses Hu&Liu's opinion lexicon for MORQA system and WordNet dictionary for addressing polarity shift problem.

The sentiment of any sentence depends on opinion word. Opinion word would express any one of the three possible

states positive, negative or neutral. We utilize bag-of-word (BOW) model to collect sentiment words from the review. Hu & Liu's opinion lexicon is used to compare and extract sentiment words from the sentences. Words that encrypt a desirable state (e.g. 'fantastic' and 'good') have a positive polarity, while words that encrypt undesirable state (e.g. 'bad' and 'worst') have a negative polarity. Normally adjectives and adverbs placed in the sentences used to express opinion polarity. Sometimes nouns and verbs also express opinions as well. In addition to that Hu &Liu's opinion lexicon, we use WordNet (Miller et al., 1990) list of synonyms and antonyms.

D. Segmentation

The product review could be simple or complex sentences. Complex sentences has more than one sentences that can express different type of sentiments. So we have to split the complex sentence into set of sub sentences. Sentence segmentation can be done by separating sentence based on commas, hyphens, semicolons, periods and conjunctions and, but etc., In this segmentation also we have to address the polarity shift problem if there are any negative sub sentences that are present in complex sentence.

E. POS Tagging

Part-Of-Speech (POS) is required to identify verb, noun, adverbs and adjectives in lexicon based on sentiment classification. The grammatical sentence are categorized into eight different POS. So our system uses this tagging to perform the desired task. POS parses each word of given text and identifies the word according to its category [3] .

The following are some rules that our model uses to extract sentiment from consumer product reviews.

1. Adjectives + Noun
2. Adverb + Adjectives + Noun
3. Noun + Adjectives
4. Noun + Verb
5. Noun + Adverb + Verb
6. Noun + Verb + Adjectives
7. Noun + Verb + Noun
8. Noun + Vern + Adverb

PROPOSED SYSTEM

We propose a system for product review sentiment analysis and classification by adding customer Q/A system to consumer review system in order to improve the accuracy of the result. The polarity shift is identified and managed with the help of opinion lexicon and antonym thesaurus in contrast of negation by considering negative reviews. Here we proposed a system which retrieves reviews form amazon online shopping portal for sentiment analysis and classification. We recommended sentence-level sentiment classification to categorize consumer's product review sentiment into positive, negative and neutral classes.

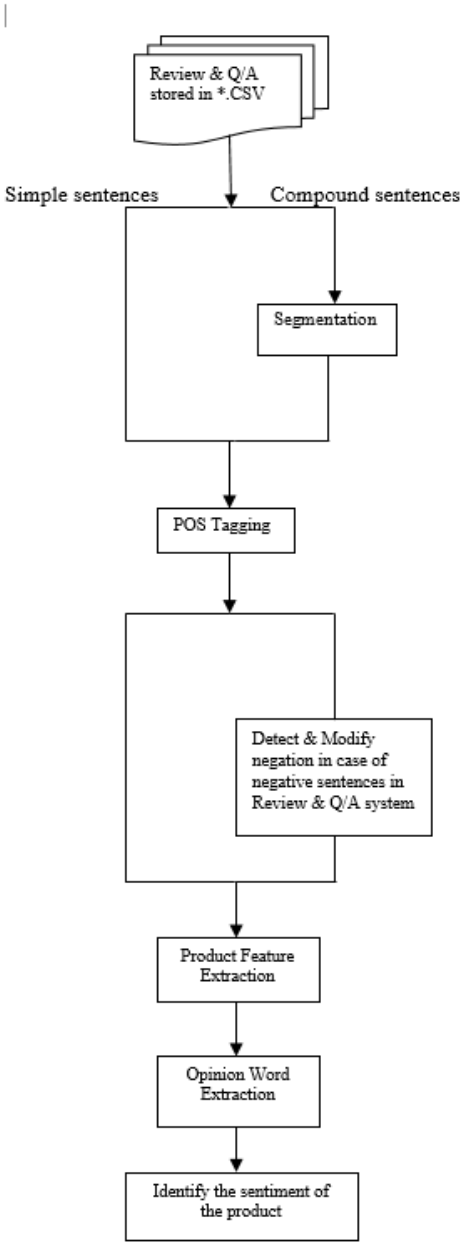

Figure 2: Flow of Proposed System

Before implementing this system, first we have to check the product review sentences are simple or complex sentences. Suppose the pulled out product reviews are simple sentences, then we implement POS tagging and train the classifier directly to find its sentiment. If the product reviews are complex sentences then segmentation and POS tagging are needed before to train the classifier to find its sentiment. Proposed system performs two additional tasks (i) Combine product review and consumer Q/A system about the product to analyse

and improve the score of the sentiment and (ii) to address the polarity shift problem in case of negation words identified in product review

A. Task 1

When we are about to buy the product through online, online reviews are our first pole which helps to understand the positive and negative sides of the product. When we are estimating a potential purchase, we may have queries about the product, e.g. "Is this lens interchangeable in Sony H300 Camera?" or "Can I insert mobile sim card for internet connection in the Micromax Canvas Lap book?" To answer these questions we have to stride through huge amount of consumer reviews to find the relevant answer. Otherwise post our question directly to the community via Q/A system. Reviews are natural and rich source of data to address the queries.

We developed a new system called MORQA- Mixture of Reviews and Questions & Answers, a novel method evaluating thousand questions (and answers) and three thousand reviews from online shopping portal. Our goal is to show sentiment analysis and classification of product features. For e.g. Opinion about Nikon D3400 camera and canon EOS 1200D camera features such as optical zoom, memory, batter life etc., we evaluate our system MORQA using manually created product feature lexicon, Hu&Liu's opinion lexicon and bigram features. MORQA system has two phases.

In first phase, we use python to evaluate the consumer Q/A system and reviews. For any product we manually create product feature lexicon, e.g. Canon EOS 1200D digital camera product feature lexicon may have battery life, memory card, flash, resolution, auto focus, lens, size, optical zoom etc., We may have to scan the questions to find whether the relevant product features exist or not. If exists, retrieve the specific product feature. Also our model check the questions to find whether the sentiment words are exist or not using Hu&Liu's WordNet lexicon .

We apply bigram feature to combine and evaluate both selected product feature and opinion words. After combining these two, check the answer part to find the Boolean answer yes or no. If yes, this specific feature positive polarity is increased by one. If no, negative polarity is increased by one. Otherwise neutral polarity is increased by one.

Algorithm

Create feature lexicon F for any product

For each question Qi in Q/A system do

Read and find any product feature Fi exist in Q_i

If exist, retrieve Fi from Qi, and scan Qi to check any

sentiword Si exist

If exist, combine Fi & Si using bigram.

check answer for each Q_i

If answer contains 'Yes' Positive Polarity is increment

by one

Else if answer contains 'No' Negative Polarity is

increment by one

Else

Neutral Polarity is increment by one.

End for

Extracted product features, opinion words and boolean answers must be stored in CSV file. This CSV file is evaluated using python graph lab create tool to make tabular representation for any product feature. People can easily understand the graphical representation than reading thousands of consumer review.

In second phase, consumer reviews are scanned to find any product feature is exist or not using product feature lexicon and similarly to find the existing opinion words using opinion lexicons. Combine these two using bigram, process and store the result in CSV file. Every two consecutive words of text are represented by using bigram features. This combination of consumer review and Q/A system will improve the sentiment score.

B. Task 2

Our proposed system addressing polarity shift problem by considering negation. For understanding consider the following example of consumer review on digital camera.

Polarity shift detection: "I don't prefer camera picture quality in iPhone 5S"

Polarity shift elimination and modification "I hate camera picture quality in iPhone 5S".

For tackling polarity shift problem, we notice that the sub sentence has sentiment word "prefer" with negation "don't" which changes the sentiment orientation from positive to negative or vice versa. Our proposed model remove the negation word "don't" and replace the sentiment word "prefer" with its antonym (hate) using WordNet. This technique removes the negation problem. The BOW model gives more reasonable solution to solve polarity shift problem using antonym dictionary.

C. Experimental Result

Our main goal is to make sure that proper results of online portal sentiments. Also we would like to save user time. We are not sustaining consumers to spend more time reading through long textual description in the reviews. So we summarize our result in the form of charts and tables. Data visualization is a

vital technology for future generation, as information is growing in size and complexity. Hence our model sum up the results as bar charts that help the consumers to outlook and analyse the sentiment of any product.

We apply this experiment on two different data sets such as Nikon D3400 digital camera and Canon EOS 1200D digital camera. In addition to that we apply this experiment to iPhone 7S+ and Micromax Canvas lap book. Table (2) represents the score of positive, negative and neutral sentiments of product features of Nikon D3400 digital camera. Figure (3) represent the cluster graph for product features opinion summary of digital camera Nikon D3400 in which we revealed the score of zoom, memory and battery features are good. So we conclude that battery, memory and optical zoom features are good in Nikon D3400.

Table 2: Nikon D3400 Features Opinion Analysis

Feature	Positive Score	Neutral Score	Negative Score
Battery	4	0	2
Memory	5	1	2
Flash	0	0	2
Auto Focus	0	0	3
Lens	3	0	0
Optical Zoom	6	1	1

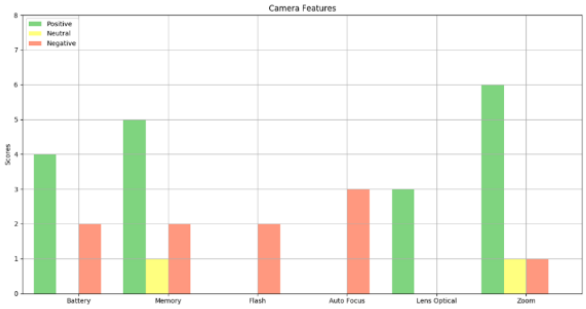

Figure 3: Cluster Graph for NIKON D3400 Digital Camera

Canon EOS 1200D digital camera product features and the counting of positive, negative and neutral sentiments of consumers are listed in the table (3). Figure (4) represent the cluster graph for product features opinion summary of digital camera Canon EOS 1200D. This shows autofocus, battery and flash features are good in canon EOS 1200D digital camera. Similarly we can produce graph for iPhone 7S+ and Micromax Canvas Lap book.

Table 3: Canon EOS 1200D Features Opinion Analysis

Feature	Positive Score	Neutral Score	Negative Score
Battery	34	12	5
Memory	20	0	2
Flash	28	17	5
Auto Focus	41	10	14
Lens	21	17	3
Optical Zoom	23	14	11

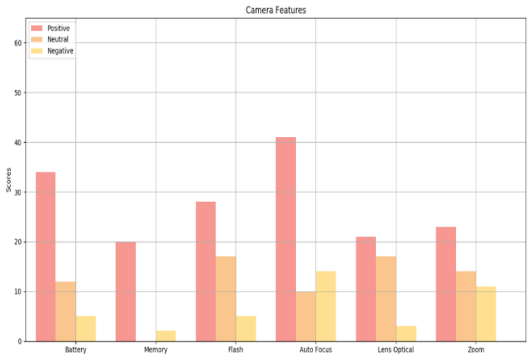

Figure 4: Cluster Graph for Canon EOS 1200D digital camera

Using our MORQA model, we can find any product feature opinion summarization. MORQA system helps to perform opinion summarization by identifying different features of any domain and evaluating number of positive, negative and neutral opinions.

CONCLUSION AND FUTURE WORK

Here we proposed system for sentiment analysis and classification of online portal reviews using product features and consumer Q/A system. Our proposed methodology classify consumers sentiment on any product based on different polarity classes such as positive, negative and neutral. The main objective of our system is to identify and resolve the negation problem in reviews to increase the sentiment score. This can be done by tackling and eliminating negation word and change the opinion word using antonym dictionary to retrieve the original sentiment of the sentence.

In future, we are planning to improve the MORQA system to identify which consumer opinions are relevant the question and

display them immediately to answer the question. i.e., to analyse which consumer opinions are 'relevant' to the question using relevance ranking and use prediction function which gives 'vote' on the answer part, if it finds the relevant consumer reviews. In future we have planned to apply dual sentiment analysis to identify and remove polarity shift problem.

REFERENCES

[1] Rui Xia, Feng Xu, Chengqing Zong, Qianmu Li, Yong Qi and Tao Li, August 2015, "Dual Sentiment Analysis: considering Two sides of one Review", IEEE transaction on knowledge and data Engineering, Vol 27, No.8.

[2] K.Zhang, W.Wu, H.Wu. Z. Li. and M.Zhou. 2014, "Question retrieval with high quality answers in community question answering". In CIKM.

[3] G.Vinodhini and RM.chandrasekaran June 2012 "Sentiment Analysis and Opinion Mining; A Survey", www.ijarcsse.com, ISSN: 2277 128X, Volume 2, Issue 6.

[4] Asst.Prof.A.Kowcika, Aditi Gupta, Karthik Sondhi, Nishit Shivhre, Raunaq Kumar, July 2013,"Sentiment Analysis for Social Media". www.ijarcsse.com , ISSN: 2277 128X, Volume 3, Issue 7.

[5] Georgios Patlogolu, Mike Thelwall, September 2013 "More than Bag-of-words: Sentence-based Document Representation for Sentiment Analysis", Proceedings of Recent Advance in Natural Language Processing, Page 546-552, Hissar, Bulgaria, 7-13.

[6] S.Li and C. Huang, 2009, "Sentiment classification considering negation and contrast transition" in Proceeding of the Pacific Asia Conference on Language, Information and Computation.

[7] Bing Liu, May 2012 "Sentiment analysis and opinion mining", Morgan & Claypool Publishers. [Book]

[8] Rui xia, Feng Xu, Jianfei Yu, Yong Qi, Erik Cambria, 2016 "Polarity shift detection Elimination and ensemble: A three-stage model for document-level sentiment analysis." Science Direct, 36-45.

[9] S.Das and M.Chen, 2001, "Yahoo! For Amazon; Extracting market sentiment from stock message boards," in Proc, Asia Pacific Finance Assoc.Annu.Conf.

[10] Minqing Hu and Bing Liu, 2004, "Mining Opinion Features in Customer Reviews", American association for artificial Intelligence.