Available online at www.elixirpublishers.com (Elixir International Journal)

Information Technology

Elixir Inform. Tech. 164 (2022) 56147-56151

ISSN: 2229-712X

A Rule-Based Approach for Sentiment Analysis of Products Review

Tapasy Rabeya and Ismail Jabiullah

Department of Computer Science and Engineering Daffodil International University Dhaka, Bangladesh

ARTICLE INFO

Article history:

Received: 12 January 2022;

Received in revised form:

17 March 2022;

Accepted: 27 March 2022;

Keywords

Sentiment Analysis,
Feature Based,
Domain Specific Sentiment,
Data Mining,
Product Review.

ABSTRACT

E-commerce has become one of the most commodious methods of shopping because of the technological revolution. Research shows internet shopping is much preferred by people rather than the traditional mode of shopping. Millions of user- generated comments are posted daily on the web, and analysis of these opinions could be more directive towards the customers and manufacturers. That makes the Sentiment analysis of online reviews one of the most sought-after research topic. This paper portrays our experimental work on domain-specific feature-based sentiment analysis of product review. In this paper, we worked with some fixed predefined core features of a product for presenting the customer's acceptance of the principal attributes of a product so that the manufacturer can improve the quality of the basic features. We have proposed a feature-oriented sentiment prediction scheme. That analyses the generated expressions from the textual reviews of a product for predicting sentiment and assigns scores for our predefined features to present a net sentiment profile of a product of all parameters. With 92% accuracy, our sentiment detection scheme is proved to be an effective way to highlight the core attributes that seem to be really beneficial to the customer and manufacturer.

I. Introduction

Online shopping has become a well-liked looking technique ever since the net has declared a takeover. The increase in technology provides sensible opportunities to the vendor to succeed in the client in a lot of quicker, easier and in economic manner. Online shopping is rising in no time in recent years [1]. According to a survey, 81% of Internet users have looked up a product online at least once. [2].

Due to the wide range of online application, analyzing the sentiment of online reviews and other user-generated content are sought-after research concern now a days. Sentiment analysis examines all comments according to various granularity, sentence level, document level, and function-based sentiment analysis.

Aspect-level or feature-based sentiment analysis performs an in depth analysis in business for analyzing their client feedback knowledge, so that they will learn extra concerning their clients and emphasis better quality products to meet their needs. In feature-based sentiment analysis, every users comment is processed to determine the polarity associated with each feature or attribute [3]. An evaluation on smartphones portrayed that, nearly 50% users of smartphone do not use their phones to their full potential [4]. But over the time as technological advancement we have been experiencing different new features of smartphone, like: numerous built-in cameras, 5G- compatible handsets, foldable gadgets, as well as holographic screen projections and an expanded AI focus. An research on smartphone owners in the US and UK aged 10-64 showed that, these fancy new features may produce the biggest buzz, however enhancements to the core phone features are actually likely to drive additional sales[5]. Every product is designed and manufactured with some basic attributes and different brand merely provide some form of differentiation against their key competitors. For example: Smartphone; there are always some common attribute that most smartphones provide, and user usually look for the features while owning a phone. Like camera, storage capacity, Speed, battery backup, display and so on.

In this paper, we worked with the core features of a product of a specific domain, like a baby diaper pail. Rather than focusing on the detailed features of a product, the work is intended to find out the customer's acceptance of the principal attributes. So that the product's company can emphasize the product's basic attributes quality. The rest of our paper is structured as follows. In section II we presents the back ground analysis of our work. Data source for our analysis is presented in section III. The methodology of our experimental work is describe in section IV. Section V presents a comparative analysis of our work with other papers that lay with the same research focus.. And Section VI describes the conclusion and finally the supporting paper references are presented.

II. Literature Review

NLP-based approach and Statistical model-based approach for creating rules are two major approaches for performing sentiment analysis. In this connection, Liu et al. First, apply text mining to extract attributes from the reviews and aggregate user reviews. Then for writing rules they used NLP approach. Here, reviews from two new apps were used: Widget apps in the Brain & Puzzle category and game apps in the Personalization category. In total they extracted six hundred textual reviews for their analysis.

SAS® Sentiment Analysis Studio 12.1 was used for their sentiment analysis, they found the SAS® Sentiment Analysis Studio 12.1 very useful as it provides the facility of defining the synonym list of products and features. They did not spell

Tele:
E-mail address: tapasyrabeya.cse@diu.edu.bd

check and no stop list was used in the study. Finally, they come up with a precision of 86 percent and 94 percent respectively on the positive and the negative directory for testing the widget app dataset. And respectively the precision of 94 percent precision and 90 percent precision on positive and negative directory for testing game app dataset. [6]

An experimental work has been done on movie reviews, a domain specific feature-based sentiment analysis. They analyzed reviews of 100 Hindi movies 10 reviews for each and reviews are flagged manually by them to evaluate the performance of the formulation of aspect-oriented algorithms. The result was satisfactory as their proposed scheme performed more exact and centered opinion profile than the straightforward document-level assumption examination [7]. Using a method based on machine learning Fang Luo, Cao et al. has done text sentiment analysis. And for textual representation, a vector space model of sentiment is used for solving the problem of lack of data. But their method is failed in word level sentiment analysis, is well performed in both document level and sentence level [8].

Sometime Sentiment analysis become very challenging task, when same word used for expressing different sentiment. A single punctuation may change a sentences sentiment. For example, "He loves this mobile?" is different from "He loves this mobile!". Different emotions can be expressed by the same sentence with different punctuation. Focused on this two research was done using Chinese review. First they proposed a new weighting algorithm and a performance-based vector model for sentiment analysis of Chinese product reviews. In the review text, they considered both the modifying relationships between words and punctuation marks and used adverbs of the degree to indicate the intensity of mood.. Lizhen et al. has proposed a feature based vector model having six tuple, specially focused on adverbs of degree and punctuations for expressing the intensity of polarity more accurately. Data sets from three different domain were used here [9] [10].

A fuzzy based product review analysis was performed by Indhuza et al. to extend the feature based classification approach which was evaluated with SFU corpus. They have collected and used 2000 user generated product comments from different websites. The general accuracy of their system was 85.58%. The representation of input data really varies the accuracy of their proposed system. For performing the classification no labeled data set for training is required for the feature-based approach. To improve the performance focus on entity recognition, co-reference resolution and domain dependency is needed [2].

A feature based sentiment analysis was done with twitter reviews. At first they select some features from the tweets and a sentiment analysis is conducted on tweets only containing those features. This proposed method avails 40% more precision compared to features extracted directly from tweets. They have done this analysis for multiple languages [11]. Another feature based analysis performed by proposing the use of SVM in new dimension as a classifier and they resolve the negation problem [12]. Another feature based analysis was done in Arabic language for extracting and weighting features and detecting sentiments from Arabic reviews [3].

A back tracking approach has been introduced for detecting sentiment of Bengali text based on a hypothesis that people most of the time express their actual emotion in the last of the sentence. For that they have proposed a backtracking algorithm and got more than 70% accuracy. Their proposed methods performance varies from the input data [13]. Another Sentiment analysis of YouTube song comment was done by using mentioned back tracking approach with slight change in the algorithm [14].

III. Data Source

We have analyzed the user's review of a specific domain: Children Diaper pail. Diaper bins are essentially trash designed specifically for storing dirty diapers. By using a diaper pail you can temporarily block the unavoidable unpleasant smell of diaper filth for several days. The data set we used in our work is a combo box of user review from 9 different domain [15]. The data was used for improving performance of a lexicon based analysis.

IV. Proposed Methodology

Our main research objective is to find the users actual sentiment of the basic features of a product. The Fig.1 shows the research methodology.

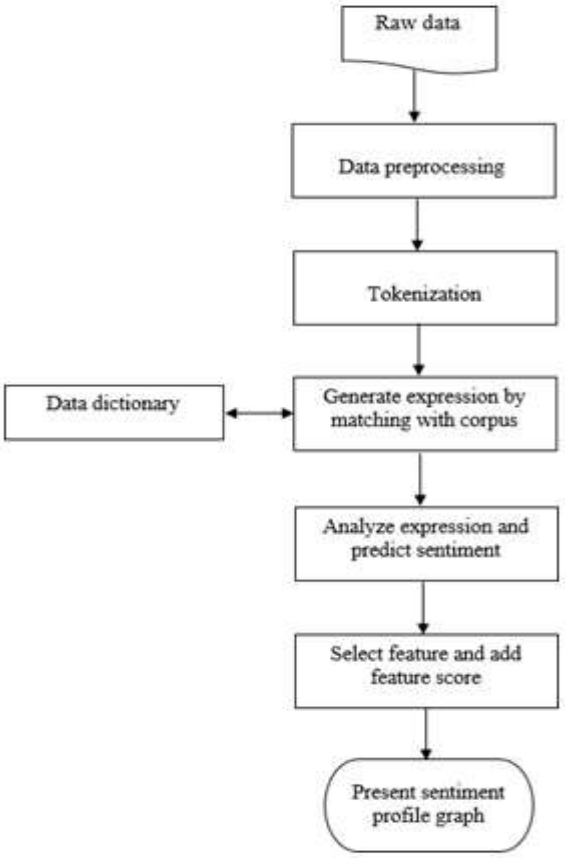

Fig. 1. Proposed methodology architecture

A. Feature Selection

For selecting the features of children diaper pail, first a general study was performed of the basic attributes for which a diaper pail is introduced. And we found some principal attributes of it. Firstly, a diaper pail should have the aptitude to block the odor from children diaper. Some pail Refills are designed for providing reliable protection from diaper odors which ordinarily tends to expensive for the user. So second attribute is the way of refilling the pail bag. Thirdly, the size and capacity of a pail. They determine how many dirty diapers it can hold, as it will determine how often it will need to be emptied. And finally the expenses of a pail use. Except these features rest of the comments will be considered under the "product's use" category. From our analysis six selected features are odor blocking, using of pail bag, size,capacity, price and product's use.

B. Data Pre-processing

For our experiment, we mainly focused on single-line comments. For long text comments, firstly we minimize the lines that don't contain any sentiment support word. Table I

demonstrates the processed user comments after removing the part containing no sentiment word. In the first comment, "We started with the Diaper Genie as most new parents do." this line is not associate with any kinds of sentiment. That's why we just kept only "We hated that thing!" this line.

Table I. Pre-Processing By Removing

Raw data	Processed data
1. We started with the Diaper Genie as most new parents do. We hated that thing!	We hated that thing!
2. The bags never fit right and then it never spun right. And after about two months it just sat there in the corner of her room and was never used.	The bags never fit right and then it never spun right.
3. We decided we had to have something so we started researching. The Diaper Champ is the best we found!	The Diaper Champ is the best we found!

And secondly, split up a long text comment into two comments. TABLE II demonstrates how we split up a long text comment. The first comment in the table is been divided into two comments.

Table II. Pre-processing by spliting up

Raw data	Processed data
1. The diaper champ couldn't be easier to use. And the biggest plus is that you don't have to buy any expensive refills.	1(a) The diaper champ couldn't be easier to use.
	1(b) The biggest plus is that you don't have to buy any expensive refills.
2. Diaper Champ is by far the best diaper pail! It holds a lot of diapers, not to mention the odor! It's easy to use (just a flip of the lid, and the diaper is gone!). It's also easy to empty out.	2(a) Diaper Champ is by far the best diaper pail!
	2(b) It holds a lot of diapers, not to mention the odor!
	2(c) It's easy to use (just a flip of the lid, and the diaper is gone!).
	2(d) It's also easy to empty out.

C. Analysis

We used four kinds of lexicon in our analysis. Positive, negative, negation and feature word. TABLE III shows the lexicon types.

Table III. Types of lexicon

Positive	P	good, great, fantastic…
Negative	N	worst, clumsy, expensive..
Negation	B	not, never, no, nothing, nobody
Feature word	F	odor, bag, size, capacity, price…

For positive and negative lexicon we have used Minqing Hu and Bing Liu's data set [16]. We have added some positive and negative word in the data set while analyzing the processed comments. And the negation and features words have been added to the dictionary for our analysis.

When an input text is given first an expression will be generated from the given text. And sentiment prediction will done by analyzing the expression. TABLE IV shows the generated expression and selected features from the sentences.

Table IV. Generated expression

Input Text	Expression	Feature
The bags never fit right and then it never spun right	BPBP	Bag
I'm very happy not to have to buy those refills.	PB	Refill
Love that you don't have to purchase expensive refills	PBN	Refill, Price
Inexpensive as baby products go!	P	Price

D. Sentiment Prediction Algorithm

Take an expression for analysis If the expression contain only P

then declare positive

else if P is followed by B and B is followed by B or none then declare positive

else if P is followed by N and N is followed none then declare positive

else if P is followed by B and B is followed by P then declare positive

else if P is followed by B and B is followed by N then declare positive

If the expression contain only N then declare negative

else if N is followed by B and B is followed by B or none then declare negative

else if N is followed by P and P is followed by none then declare positive

else if N is followed by B and B is followed by N then declare negative

else if N is followed by B and B is followed by P then declare negative

If the expression contain only B then declare negative

else if B is followed by P then declare negative

else if B is followed by N then declare negative

else if a string is long

then find the mentioned string above to reform a string to find the result

E. Generating Feature Score

For receiving sentiment profile of a product, our proposed aspect-level sentiment analysis algorithm is a unique way. The Resulting sentiment profile is extremely useful for users, informative and easy to understand. Once sentiment of a text is predicted then it will score a feature. For every feature there are two types of scoring. Positive scoring and negative scoring. TABLE V presents the table representation of Sentiment Profile of Champ Diaper Pail and Fig.2. Presents the graphical representation.

Table V. Sentiment profile of champ diaper pail

Features	Number of Positive comments	Number of Negative comments
Odor	63	51
Bag refill	57	27
Size	0	3
Capacity	3	0
Price	27	0
Product's use	105	45

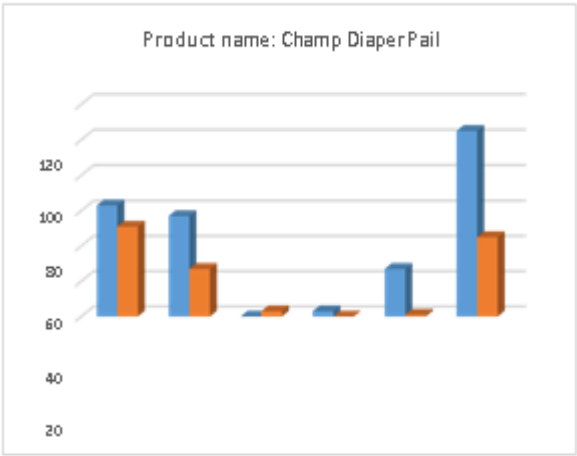

Fig. 2. Sentiment Profile of Champ Diaper Pail

F.Result Analysis

Our fixed featured based analysis on specific domain is done using 354 user generated product reviews. We have used 4 types of lexicon for generating expression for

sentiment detection purpose. Among 354 comments our algorithm perform 282 right prediction and 36 wrong prediction. For some comments our proposed algorithm failed to predict any sentiments. As our system deals with generated expression, so if a comment does not contain any positive, negative or negation word, our system failed to predict any sentiment. But sometimes the comments depict very strong sentiment. For example-"Holds a LOT of diapers--a full week of toddler diapers!" This comment strongly reveal the capacity of the Diaper pail but our system fail to predict anything though it could not generate any expression from it. TABLE VI & TABLE VII our evaluation result.

Table VI. Sentiment prediction result

Total Comments	Right Prediction	Wrong prediction	Failed to predict
354	282	36	36

Table VII. Number of true positive, true negative, false positive and false negative

TP	TN	FP	FN
207	75	6	30

Here, TP is true positive which indicating the number of correctly classified sentences as positive, FP is false positive which indicating the number of incorrectly classified sentences as positive, TN is true negative which indicating the number of correctly classified sentences as negative and FN is false negative which indicating the number of incorrectly classified sentences as negative.

Recall, precision and F-measure are used for evaluating the performance of our algorithm, which are defined as follows:

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recal} \quad (3)$$

With 97.2% precision and 87.3% recall, our Empirical results shows 92% accuracy of our proposed algorithms performance. Empirical results indicate that our sentiment detection scheme an effective ways to highlight the core features that are the most relevant to the customer and manufacturer. The performance of the algorithm varies from the input data representation.

V. Comparative Analysis

All the three papers mentioned below are concentrated on feature-based sentiment analysis. TABLE VIII presents a comparative analysis of our work with three other papers that lay with the same research focus.

Table VIII. Comparative analysis

Attributes	Paper 1	Paper 2	Paper 3	Proposed method
Analyze with Fixed features	NO	NO	NO	YES
Use POS tagging for Feature Extraction	YES	YES	NO	NO
Expression generation form text	NO	NO	NO	YES
Use new approach to predict sentiment	YES	NO	YES	YES
Use of precision and recall	YES	YES	YES	YES

Paper 1 *Mining Opinion Features in Customer Reviews [17]. Paper 2 *Sentiment analysis of movie reviews: A new feature- based heuristic for aspect-level sentiment classification [7].

Paper 3 *Feature-based Sentiment Analysis Approach for Product Reviews [10].

All the three papers mentioned below are concentrated on feature-based sentiment analysis. TABLE VIII presents a comparative analysis of our work with three other papers that lay with the same research focus.

In our paper, we worked with some core features of a product unless analyzed all the features of a product. For that, background analysis was required to determine the fixed features. On the other hand, Paper 1,2 & 3 have not worked with fixed features. Paper 1 had identified two kinds of features from analyzing users' comments: the most frequent and infrequent features. For identifying the attribute in different film awards, movie review sites, and film magazines and working out a list of aspects the Paper 2 has made an extensive search. Paper 3 has used a feature-based six-tuple vector model which contains both features and opinions that describe reviewers' opinions on the features.

We have not used POS tagging for feature extraction. As we analyzed with fixed features, so that we have added some feature words in our corpus. We have four kinds of words in our corpus: positive, negative, negation, and feature word. Paper 1 and 2 have used POS tagging for extracting features from the user comments. Paper 1 additionally used a human tagger to evaluate the discovered features, by reading manually all the reviews and producing a manual feature list for each product. Paper 3 has proposed an algorithm for extracting feature words from a sentence.

For analyzing sentiment we had an intermediate step called generating expression. We generate expression for every sentence. We have denoted our words with a single alphabet to generate expression. Like- P for Positive word, N for negative word, and B for negation word.

Example: BPBP is the expression of "The bags never fit right and then it never spun right."

The bags never fit right and then it never spun right

B P B P

Finally, sentiments are predicted by analyzing these expressions. Paper 1, 2 & 3 don't generate any kinds of expression for sentiment analysis.

We have used a new and simple rule-based approach for predicting the sentiment of a comment. To determine every sentence sentiment polarity a SentiWordNet based approach is used by Paper 2, on the other side a new weighting algorithm is proposed for analyzing sentiment of Chinese reviews by paper 3, and paper 1 used their proposed algorithm.

All of the papers used precision and recall for the performance calculation of their empirical work.

VI. Conclusion

Sentiment analysis is complicated and a quite challenging task as the expression in natural language changes in no time, no method can exhaustively deal with every situation. In our work we focused on fixed features of a product so that user can only get a clear idea about the basic attributes from the products comment. To do so, an experiment is performed with a generic proposed algorithm. In this algorithm, for sentiment analysis first the system generates expression of the comments with three kinds of word from our dictionary and identify the feature word from the text. After applying our sentiment prediction algorithm for every predefined features, number of positive and

negative comments are calculated to depict the sentiment profile graph. With 92% accuracy the algorithm used for feature based sentiment analysis is very simple and no previous training is required. It can be used to produce very detailed and useful sentiment profile of basic features of any product.

References

[1] Lopes, Priyanka Nazreth, Tanaya Prabhutendulkar, and Deepa Yogesh Naik. "A Comparative Study On Online v/s Offline Shopping of Eyewear." Reflections-Journal of Management 5 (2015).

[2] Indhuja, K., and Raj PC Reghu. "Fuzzy logic based sentiment analysis of product review documents." In 2014 First International Conference on Computational Systems and Communications (ICCSC), pp. 18-22. IEEE, 2014.K. Elissa, "Title of paper if known," unpublished.

[3] Abd-Elhamid, Laila, Doaa Elzanfaly, and Ahmed Sharaf Eldin. "Feature- based sentiment analysis in online Arabic reviews." In 2016 11th International Conference on Computer Engineering & Systems (ICCES), pp. 260-265. IEEE, 2016.Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[4] Public policy Anlysis, available at <<https://blogs.oregonstate.edu/econ439/2014/01/17/smartphones-come-many-features-majority-used/>> last accessed on 07-01-2022 at 5:03 PM.

[5] GWI, available at<<https://blog.gwi.com/chart-of-the-week/smartphone- features-consumers/>>last accessed on 07-01-2022 at 5:03 PM.

[6] Liu, Jiawen, Mantosh Kumar Sarkar, and Goutam Chakraborty. "Feature- based Sentiment analysis on android app reviews using SAS® text miner and SAS® sentiment analysis studio." In Proceedings of the SAS Global Forum 2013 Conference, vol. 250. 2013.

[7] Singh, Vivek Kumar, Rajesh Piryani, Ashraf Uddin, and Pranav Waila. "Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification." In 2013 International Mutli- Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), pp. 712-717. IEEE, 2013.

[8] Luo, Fang, Cheng Li, and Zehui Cao. "Affective-feature-based sentiment analysis using SVM classifier." In 2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD), pp. 276-281. IEEE, 2016.

[9] Lizhen, Liu, Song Wei, Wang Hanshi, Li Chuchu, and Lu Jingli. "A novel feature-based method for sentiment analysis of Chinese product reviews." China communications 11, no. 3 (2014): 154-164.

[10]Wang, Hanshi, Lizhen Liu, Wei Song, and Jingli Lu. "Feature-based Sentiment Analysis Approach for Product Reviews." J. Softw. 9, no. 2 (2014): 274-279.

[11] Erdmann, Maike, Kazushi Ikeda, Hiromi Ishizaki, Gen Hattori, and Yasuhiro Takishima. "Feature based sentiment analysis of tweets in multiple languages." In International Conference on Web Information Systems Engineering, pp. 109-124. Springer, Cham, 2014.

[12] Devi, DV Nagarjuna, Chinta Kishore Kumar, and Siriki Prasad. "A feature based approach for sentiment analysis by using support vector machine." In 2016 IEEE 6th International Conference on Advanced Computing (IACC), pp. 3-8. IEEE, 2016.

[13] Rabeya, Tapasy, Sanjida Ferdous, Himel Suhita Ali, and Narayan Ranjan Chakraborty. "A survey on emotion detection: A lexicon based backtracking approach for detecting emotion from Bengali text." In 2017 20th international conference of computer and information technology (ICCIT), pp. 1-7. IEEE, 2017.

[14] Rabeya, Tapasy, Narayan Ranjan Chakraborty, Sanjida Ferdous, Manoranjan Dash, and Ahmed Al Marouf. "Sentiment analysis of bangla song review-a lexicon based backtracking approach." In 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), pp. 1-7. IEEE, 2019.

[15] From, G. M. R. K. B. "Jump to: navigation, search (Ding et al., 2008)⇒ Xiaowen Ding, Bing Liu, and Philip S. Yu.(2008)."A Holistic Lexicon- based Approach to Opinion Mining." In: Proceedings of the International Conference on Web Search and Web Data Mining (WSDM 2008). doi: 10.1145/1341531.1341561 Subject Headings: Opinion Mining, Sentiment Analysis, Context-dependent Word, Product Mention,Lexicon-based Algorithm."2

[16] Hu, Minqing, and Bing Liu. "Mining and summarizing customer reviews." In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 168-177. 2004.

[17] Hu, Minqing, and Bing Liu. "Mining opinion features in customer reviews." In AAAI, vol. 4, no. 4, pp. 755-760. 2004.