

Aspect-Based Sentiment Words and Their Polarities Using Chi-Square Test

Pradnya Bhagat¹, Pratik D. Korkankar^{1,2},
Jyoti D. Pawar¹

¹ Goa University,
Goa Business School,
India

² Dnyanprassarak Mandal's,
College and Research Centre, Goa,
India

{dcst.pradanya, dcst.pratik, jdp}@unigoa.ac.in

Abstract. Most of the user-preferred products on e-commerce websites are accompanied by a massive number of product reviews and manually analyzing each review to understand the features and the user opinions associated with the products is an inconceivable task. A single domain of products can contain thousands of different products and an equally significant number of associated product features/aspects, thereby making the polarity of the sentiment words in the product reviews vary widely according to the aspect with which they are associated. The paper uses the Chi-square Test statistical measure to automatically calculate the aspect-based polarity of sentiment words in a given domain. The results of the method are tested on two different domains. The experimental results show that the method delivers an accuracy of more than 75% in both domains. The method also helps in discovering strong domain-specific polar adjectives that might be missing in universal sentiment lexicons.

Keywords. Aspect/feature words, sentiment words, polar words, universal sentiment lexicon, chi-square test.

1 Introduction

In this age of e-businesses where more and more people are turning towards e-commerce websites to fulfill their day-to-day needs, writing and sharing of product reviews has become an imperative part of our online transactions.

User reviews help the users to share with other users their views and experiences about various products, in turn benefiting other prospective buyers of such products.

Also, user reviews are written by users of the products rather than the brands and hence are considered more credible and trustworthy by other users [17, 27].

Moreover, since these user reviews are publicly shared, companies too have to make an effort to deliver better quality products and services to the users to avoid negative publicity.

User reviews also help e-commerce websites understand the various features of different products and their quality, thereby helping them make an informed choice about the products displayed and recommended to their customers.

Every product on e-commerce websites is accompanied by a massive number of reviews and manually reading every review to understand the features and sentiments associated with them is an almost impossible task.

Hence, companies take the help of sentiment analysis techniques to make sense from such reviews and understand the features and the sentiments expressed.

Sentiment Analysis (SA) techniques automate and help us identify, analyze and summarize the features and the corresponding sentiments

Algorithm 1 Identifying the aspect-based polarity of sentiment words using Chi-square test

Input: 1) Domain Specific Corpus Tagged with POS.
2) Feature words identified.

Output: Sentiment-Feature pairs

```

1: for Each feature word F in the corpus do
2:   for Each sentiment word S associated with F do
3:     T:=Total Count ( $S_F$ )
4:     P:=Count of ( $S_F$  in Positive Reviews)
5:     N:=Count of ( $S_F$  in Negative Reviews)
6:     Expected Count:=  $T/2$ 
7:     if Expected Count > 5 then
8:        $\text{Chi}^2(S_F) = ((P - \text{ExpectedCount})^2 +$ 
9:          $(N - \text{ExpectedCount})^2) / \text{ExpectedCount}$ 
10:      if  $\text{Chi}^2(S_F) > \text{Threshold}$  then
11:        if  $(P - N) > 0$  then
12:          Polarity( $S_F$ ) = Positive
13:        else
14:          Polarity( $S_F$ ) = Negative
15:      else
16:        Continue to next  $S_F$ 
17:    else
18:      Continue to next  $S_F$ 

```

Table 1. Domain Wise correctly retrieved pairs

	Positive	Negative
Grocery	286	52
Cell phones its related accessories	268	55

associated with the features in the massive amount of textual content that is available to the users on the Internet [19, 20]. Traditional SA techniques focus on the use of universal sentiment lexicons [11, 14] to understand the polarity of the sentiment words extracted.

Universal sentiment lexicons are basically compilations of sentiment words divided into positive, negative, and neutral categories which help to identify the polarity of the sentiment words that have the same or universal polarity across all domains [13].

Some examples of such words are good, bad, great, worst, which irrespective of the domain being used will always convey the exact same polarity assigned to them by the sentiment lexicon. But the traditional methods fail to identify the correct polarity of the sentiment words where the polarity is dependent on the domain being used.

A classic example of such a word is unpredictable, where the word is proven to have a positive polarity in the movie domain but negative polarity in the car domain [22]. As a result, many approaches to SA have been proposed by taking into account the domain-dependent polarity of the sentiment words.

But in the real world, the polarity of the sentiment words may differ even within the same domain with respect to the feature word with which it is associated. For instance, if we take the food domain, the word 'cold' is interpreted as positive when we say 'the milkshake is cold' but when we say 'the tea served is cold' the same word expressed may get assigned a negative polarity.

Hence, there is a need to identify the aspect or feature-based polarity of the sentiment words. In this research paper, we use the terms product features and product aspects alternatively, and they both mean the same.

The research experiment demonstrated in this paper is inspired from [22], which attempts to detect the domain-specific polarity of sentiment words, but fails to look at the change in sentiments at the granularity of aspects.

Our research attempts to calculate the polarity of sentiment words at the granularity of aspects within the context of the same domain. The proposed work attempts to find the aspects and the aspect-based polarity of sentiment words in user reviews without the use of a sentiment lexicon.

The first part of the paper deals with extracting the feature words occurring in user reviews. Nouns are the words that usually contribute to feature words in user reviews. But many times, other non-feature nouns too can occur in that dataset which may not be a part of the feature nouns set.

Hence in order to extract distinguishing feature nouns from non-feature nouns, we proceed with an assumption that feature nouns occur in close proximity to sentiment words [8].

As a result, we consider only the nouns that are associated with a sentiment word as feature nouns. The second part of the study deals with finding the sentiment words and the aspect-based polarity of those sentiment words associated with the aspects. Sentiment words usually occur as adjectives in the dataset.

Hence, we extract all the identified feature nouns and their corresponding adjectives. Chi-square Test statistical measure [23] is employed to calculate the polarity of adjectives with respect to the corresponding aspects.

The Chi-square test basically looks at the difference in the expected and observed occurrence frequencies of the sentiment word in association with a particular feature word in positive and negative reviews.

We proceed with the assumption that features are uniformly distributed across positive and negative reviews in a dataset. These features will have associated sentiment words with them.

If a sentiment word can have both positive or negative polarities, without having any background knowledge about the distribution of the word, we must expect a sentiment word in association with a feature word to occur in both positive and negative reviews uniformly.

Instead, if a sentiment word in association with a feature word has a tendency to occur significantly in any one category of reviews, we can consider that sentiment word to be polar with respect to that particular feature, and we can assign the corresponding polarity to the sentiment word.

But in some cases, the sentiment word with respect to a feature word may occur in any one category just by chance and the frequency of occurrence may not be sufficient to safely tag the sentiment word with respect to a feature word as polar with the corresponding polarity.

Hence, we use the Chi-square test to take into account this magnitude of the difference in the occurrence frequencies of sentiment word - feature noun (sentiment-feature) pairs and reject the ones with non justifiable differences with the help of a threshold.

If the word satisfies the Chi-square test, it indicates that there is a significant difference between the expected and observed count of the word in the positive and negative reviews with respect to the feature and we can assign a aspect-based polarity to this particular word.

The remainder of the paper is organized as follows. Section 2 describes the related work studied. Section 3 is titled proposed methodology and explains the adaptation of the Chi-square test

Table 2. Examples of Sentiment-Feature pairs retrieved from grocery domain

Sentiment	Feature	Polarity
Positive		
affordable	price	Positive
excellent	flavor	Positive
great	quality	Positive
Negative		
high	price	Negative
artificial	flavor	Negative
low	quality	Negative

Table 3. Examples of Sentiment-Feature pairs retrieved from cell phones and its related accessories domain

Sentiment	Feature	Polarity
Positive		
amazing	phone	Positive
big	screen	Positive
durable	case	Positive
Negative		
cheap	phone	Negative
broken	screen	Negative
bad	case	Negative

to calculate the aspect-based polarity of sentiment words for a specific domain. Section 4 explains the implementation details and the datasets used and Section 5 elaborates on the experimental details. Section 6 presents the results and discussions, and finally, Section 7 states the conclusion.

2 Related Work

Various techniques have been employed in the literature to compute the polarity of sentiment words according to the context of the domain or associated features.

[12] discusses a probabilistic rating framework that calculates the sentiment orientations (SO) and strength of the opinion words using a relative-frequency-based method.

Table 4. Domain-wise Accuracy obtained using the Chi-Square Test

	Grocery	Cell phones
Accuracy	78.78%	75.64%

Table 5. Sentiment words changing their polarities within the same domain according to the associated features (Grocery and Cell phones & its related accessories domain)

Sentiment	Feature	Polarity
Grocery domain		
few	calories	positive
	pieces	negative
high	quality	positive
	price	negative
little	oil	positive
	pieces	negative
low	price	positive
	quality	negative
Cell Phones and its related accessories domain		
cheap	protector	positive
	quality	negative
extra	protection	positive
	money	negative
high	speed	positive
	price	negative
more	space	positive
	money	negative

The method allows semantically similar words to have different SO, thereby overcoming the limitations of a traditional SA.

The method further extends as a rating inference model and transforms user preferences expressed as unstructured, natural language texts into scalar ratings that can be used to perform Collaborative Filtering (CF) tasks.

[6] proposes to convert unstructured reviews into rich product descriptions and then uses the generated product descriptions to build a recommender system.

The work presents a recommendation ranking strategy that combines similarity and sentiment words to suggest products that are similar or

better than the requirements of the user according to the opinion of reviewers for a particular product. [9] describes how product features can be automatically mined from the user reviews and how these features can be aggregated at the product level to product cases.

Next, the work explains how these product features can be associated with sentiment information to reflect the opinions of reviewers, whether positive, negative, or neutral.

In continuation to this work, [7] explores the use of clustering techniques to identify common features that might be treated as independent features using any other standard opinion mining approach.

[22] applies the Chi-square test statistical measure to detect the polarity of the sentiment words based on the difference in their counts in the positive and negative reviews. The work addresses the polarity difference of words at the domain level but does not attempt to study the polarity change that might be visible at the feature level.

[16] proposes a method to automatically build context-sensitive, domain-specific sentiment lexicons using user ratings as emotional signals. The paper takes into account the change in polarity of sentiment words with respect to the associated features to assign a score to each sentiment word, which helps to find the intensity and the polarity of the sentiment word with respect to that particular feature.

But the work fails to filter out the words that might be occurring in either of the categories just by chance and there might not be any significant differences in their occurrences in positive and negative reviews.

[1] presents a comparative study of various feature extraction methods used in literature to extract features from user reviews. [2] proposes a method to extract the aspect-based nouns from the literature survey using a method based on the Latent Dirichlet Algorithm (LDA) [3].

The paper also proposes a method to find the polarity of sentiment words based on their occurrences in reviews of different polarities. The polarity of the sentiments calculated is domain-specific and not aspect-based.

[5] proposes a method to extract implicit aspects from opinionated documents using Conditional Random Fields [26].

3 Proposed Methodology

The proposed work attempts to find the aspects and the aspect-based polarity of sentiment words in user reviews without the use of a sentiment lexicon.

We first attempt to extract the aspects from the user reviews, and then we attempt to find the associated sentiment words and the aspect-based polarity of those sentiment words using the Chi-square test statistical measure.

We POS tag [24, 25] the reviews and consider only nouns as the part of speech that contribute to product features. But not all nouns can contribute to product features, and we consider only the nouns that occur in close association with adjectives as feature nouns.

The assumption is based on the observation that since the person is interested in describing his/her experience with using a particular product or its feature, he/she will be expressing their sentiments about a product or a feature in the reviews; making use of adjectives.

But other randomly occurring nouns will not have such sentiment words associated with them. Hence, we consider only those nouns accompanied by adjectives as feature nouns. For example, if we have a sentence from the cell phones domain:

My friend advised me to buy this awesome mobile because it has this stunning look and attractive features [1]. POS tagging of the above sentence would give us:

My_PRP friend_NN advised_VBD me_PRP to_TO buy_VB this_DT awesome_JJ mobile_NN because_IN it_PRP has_VBZ this_DT stunning_JJ look_NN and_CC attractive_JJ features_NNS

The nouns occurring in the above sentence are friend, mobile, look, and features. Out of these, the nouns we would be interested in are look, mobile and features since they belong to the domain of cell phones.

Table 6. Number of strong polar words obtained in the Grocery and Cell phones & related accessories domains

	Strong Positive	Strong Negative
Grocery	92	18
Cell phones related accessories	64	10

Table 7. Kappa Score for inter-rater agreement

Domain	Grocery	Cell phones and its related accessories
Cohen's Kappa Coefficient	0.5679	0.5205

As seen, the nouns mobile, look and features have some adjectives associated with them since the users want to express their opinions about the features but the noun friend does not have any adjectives associated with it.

As it is not a feature related to mobile phones, the user is not interested in expressing an opinion on it in a review post. We also filter out the randomly occurring feature nouns that occur very sparsely in the dataset by considering only the nouns that appear above a particular threshold.

We divide the reviews into positive and negative categories based on the number of star rating associated with the review text. For negative reviews, we consider all the reviews with a rating of 1 star, as 1 star is the minimum rating that could be given on Amazon e-commerce website. For positive reviews, we consider all the reviews having a 5 star rating.

Next, we consider the adjectives occurring with the feature nouns as the sentiment word associated with that particular feature.

We extract all such sentiment-feature pairs which are required for our further computations.

We proceed with a null hypothesis that; if an adjective is neutral with respect to a feature and doesn't have any positive or negative sentiment associated with it, then it should occur an almost equal number of times in association with that particular feature word in positive and negative reviews.

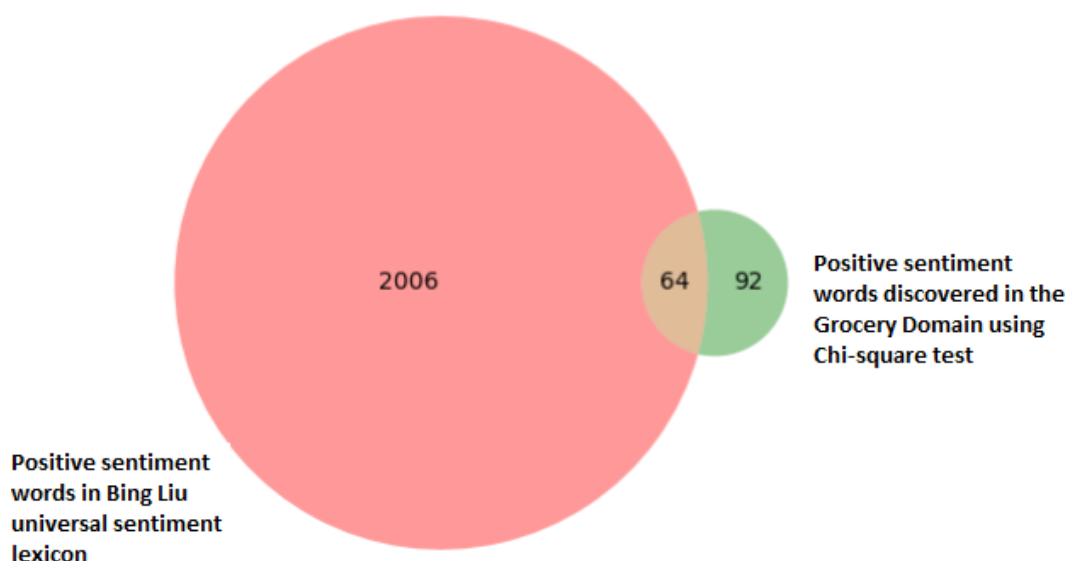


Fig. 1. Intersection of the strong positive sentiments words from the grocery domain with the Bing Liu sentiment Lexicon (Positive)

An example of such a sentiment- feature pair is front camera; although front is an adjective, there is no sentiment associated with it and the pair will occur uniformly in positive and negative reviews with corresponding positive or negative adjectives.

If we observe that a particular sentiment- feature pair occur significantly more number of times in a particular category of reviews than the other category of reviews, we reject the null hypothesis and label the sentiment word as polar with respect to that feature word according to the category of reviews it is majorly occurring in.

For example, the sentiment-feature pair awesome camera will most probably occur majorly in positive reviews than in the negative reviews, so we can consider the word awesome positive with respect to the feature camera.

But there may also be some sentiment-feature pairs which may occur in any one of the categories just by chance and may not be real representation of their polarity.

We use Chi-square test statistical measure to find out if an adjective word actually has a tendency of leaning towards a particular sentiment category than the other with respect to a particular feature.

Chi-square test also helps us in deciding whether the occurrence of a particular word with a particular feature in a particular category of reviews is significant or are we seeing it by chance.

If a word is categorised as significant by Chi-square test by resulting in a Chi-square value that is more than the threshold set by us, we can safely state that the occurrence of the word in a particular category of reviews is not by chance, but because of the domain specific polarity of the adjective with respect to a particular feature, which makes the word more frequent in one of the positive or negative categories of reviews.

4 Implementation Details and Datasets Used

The experiment is implemented using Python Programming Language [21]. The text processing tasks are carried out using the Natural Language Toolkit (NLTK) Library[15].

The dataset used by us is a collection of reviews from Amazon.com [10, 18]. We test our experiment on 2 distinct domains of data namely, Grocery and Cell Phones & its related accessories.

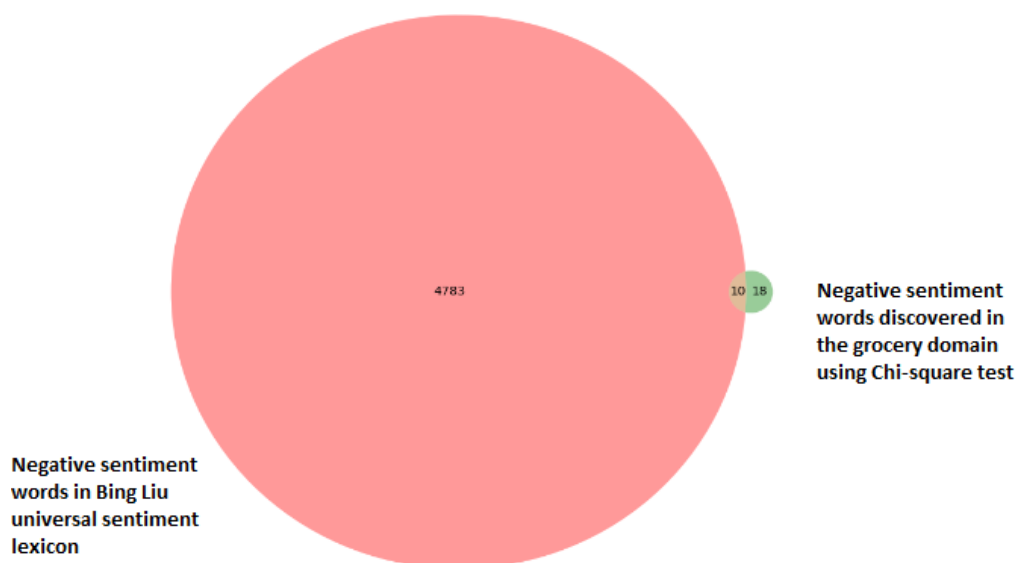


Fig. 2. Intersection of the strong negative sentiments words from the grocery domain with the Bing Liu sentiment Lexicon (Negative)

5 Experimentation Details

The two domains considered for experimentation in our study are 1) Grocery and 2) Cell Phones and its related accessories. We have considered 10,000 reviews in each of these domains, which are further divided into 5,000 positive reviews and 5,000 negative reviews.

To extract the feature words, we extract all the nouns that are associated with adjectives conveying positive or negative sentiment.

We use the universal sentiment lexicon published by Bing Liu [11, 14] to identify if the adjective with which the feature noun is associated is a sentiment word or not.

Universal sentiment lexicons only contain the sentiment words which are strongly polar, independent of the domain being used in.

The purpose of this step is only to identify feature nouns from the non feature nouns based on our observation that feature nouns will mostly have sentiment words associated with them to describe the users' opinion.

We do not need to consider the polarity of the sentiment words for this.

As a result, use of a universal sentiment lexicon is sufficient at this stage to identify feature nouns that are associated with strong polar words.

Next, we sort these feature nouns in the decreasing order of their occurrence frequency and keep only those that have their occurrence frequency above a particular threshold.

We select the threshold as 5 in our experiment, since experimentally we observe that features occurring less than 5 times are mostly randomly occurring features and do not add any value to our results.

Next, we take every identified feature word and find all the adjective words associated with it. We call these pairs sentiment-feature pairs.

We calculate the Chi-square value of every sentiment-feature pair by taking into account the expected count and the actual observed count of the sentiment word with respect to a feature in positive and negative reviews. Algorithm 1 summarizes the process.

We request two human annotators to evaluate the results obtained through our experiment. Next, we use Cohen's Kappa inter-rater agreement [4] to validate the evaluation done by both the judges independently.

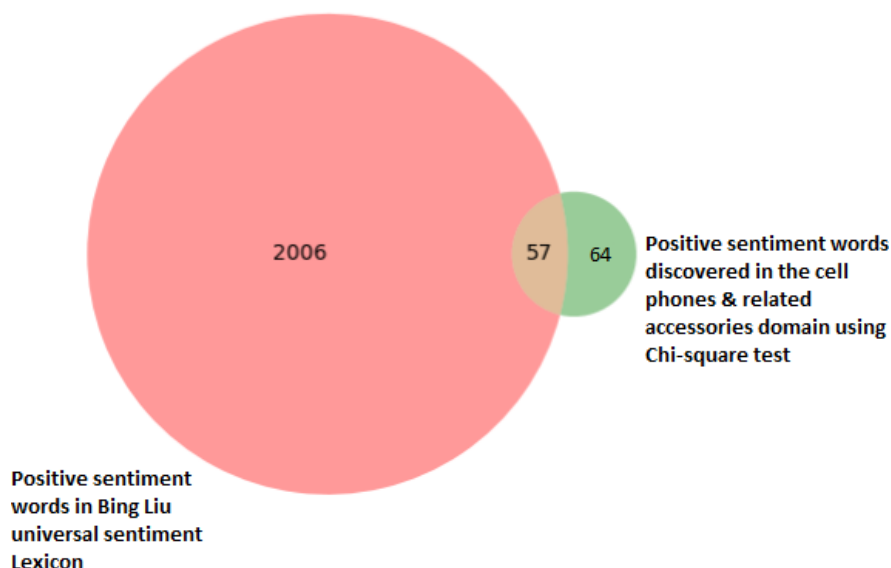


Fig. 3. Intersection of the strong positive sentiments words from the cell phones and related accessories domain with the Bing Liu sentiment Lexicon (Positive)

6 Results and Discussions

As shown in Table 1, using the proposed method we obtain 286 and 52 sentiment-feature pairs correctly classified as positive and negative respectively for the grocery domain.

Cell phones and its related accessories domain gave us 268 and 55 positive and negative sentiment-feature pairs respectively.

We can see that the number of sentiment-feature pairs expressing positive sentiments is higher than the sentiment-feature pairs expressing negative sentiments for both the domains.

We finalize the Chi-square test threshold value as 1.07 after experimentation, as we observe that a smaller value wrongly categorizes too many values as polar and a higher value significantly categorize polar words as non-polar, thereby affecting the recall of the system and missing out on important sentiment-feature pairs.

We observe that with a threshold of 1.07, we were even able to classify boundary words; words that do not have a great amount of difference in their distribution in positive and negative documents but are still significant.

Table 2 and Table 3 show some of the pairs retrieved for the grocery and cell phones & its related accessories domain respectively using the Chi-square test. The accuracy of the results obtained is 78.78% and 75.64% for the grocery and cell phones domain respectively.

Table 4 summarizes these results. Grocery domain is generally easier to understand than cell phones domain which may have a lot of technical terminology.

As a result, people find it easier to write the reviews for grocery and also for the annotators in our experiment, it was easy to understand the sentiment-feature pairs.

Table 5 lists some of the words that displayed opposite polarities according to the features being associated with in the same domain, and hence it proves our initial premise that the polarity of sentiment words may even change within the same domain with respect to the associated feature words.

We also made a list of words that always show a particular polarity in the domain irrespective of the feature with which they are associated. We call these words strong domain-specific polar words.

Table 6 shows us the number of strong polar words obtained for both the domains. As we can see, for the grocery domain, we encountered around 92 positive and 18 negative polarity words that did not change their polarities irrespective of the feature with which they were associated.

When compared with the Bing Liu universal lexicon, we obtained around 28 words that were strongly positive in the grocery domain but were missing from the universal sentiment lexicon.

Some of the examples of these words include Almond, balsamic, Chinese, coconut, creamy, daily, dark, deep, digestive, entire, extra, french, full, goji, green, hemp, Himalayan, instant, Light, local, mild, mixed, natural, nutritional, nutty, olive, orange, organic, plain, quick, raw, regular, resealable, etc. We see that, these are the words people usually prefer in connection with the food/grocery items.

Similarly, we obtained around 18 strong negative adjectives in the grocery domain out of which around 8 were not present in the Bing Liu universal sentiment lexicon. Some of the examples include artificial, chemical, metallic, plastic, etc.

The word artificial when used in the food or grocery domain is bound to derive negative polarity. People hardly speak positively about artificial colors/artificial flavors.

But the word need not be negative in other domains. As a result, the word is absent from the universal sentiment lexicon.

We received around 64 words in the Cell Phones domain that showed positive polarity irrespective of the feature being used with and 10 negative words that always showed a negative polarity irrespective of the feature being used with.

We compared the results with the Bing Liu sentiment lexicon and we did find strong positive words in the cell phones & related accessories domain that were missing from the universal sentiment lexicon; some of these words include Android, ballistic, black, blue, expensive, extended, external, larger, light, little, long, magnetic, pink.

All the strong negative words discovered in the cell phone domain using the Chi-square test were a part of the Bing Liu sentiment lexicon too.

The accuracy of the annotation is verified with the help of Cohen's Kappa inter-rater agreement.

Table 7 shows the inter-rater agreement achieved on analyzing the annotations done by 2 human judges independently.

Since, the Kappa score value is more than 0.5 in both cases we can conclude that, there is an agreement between the annotations given by both the judges, and hence the judging done is considered as valid.

7 Conclusion

The research work attempts to calculate the domain-specific sentiments of the adjectives with respect to the corresponding features/aspects in user reviews.

The experiment is tested on two different domains and the results show that the method is able to calculate the polarity of unique sentiment-feature pairs with an accuracy of more than 75% in both the domains tested.

The method was also able to find strong domain-specific polar words that were missing from the universal sentiment lexicons.

We were even able to identify sentiment words that changed their polarity based on the feature words associated with them within the same domain.

The future work consists of conducting a live user trial by providing the sentiment-feature pairs generated using the proposed method as recommendations to the users while writing reviews for products in a particular domain.

This can help us measure the difference between the quality of reviews written without any assistance and the reviews written with the help of recommendations. The work can even be extended to generate structured summaries from user reviews for different products.

Acknowledgments

This publication is an outcome of the research work supported by Goa University and Visvesvaraya PhD Scheme, MeitY, Govt. of India (VISPHD-MEITY-2002).

References

1. **Bhagat, P., Pawar, J. D. (2018).** A comparative study of feature extraction methods from user reviews for recommender systems. *Proceedings of the ACM India joint international conference on data science and management of data*, pp. 325–328. DOI: 10.1145/3152494.3167982.
2. **Bhagat, P., Pawar, J. D. (2021).** A two-phase approach using LDA for effective domain-specific tweets conveying sentiments. *Computational Intelligence and Machine Learning: Proceedings of the 7th International Conference on Advanced Computing, Networking, and Informatics (ICACNI 2019)*, pp. 79–86. DOI: 10.1007/978-981-15-8610-1_9.
3. **Blei, D. M., Ng, A. Y., Jordan, M. I. (2003).** Latent dirichlet allocation. *Journal of machine Learning research*, Vol. 3, pp. 993–1022.
4. **Cohen, J. (1960).** Introduces kappa as a way of calculating inter rater agreement between two raters. *Educational and Psychological Measurement*.
5. **Cruz, I., Gelbukh, A. F., Sidorov, G. (2014).** Implicit aspect indicator extraction for aspect based opinion mining. *International Journal of Computational Linguistics and Applications*, Vol. 5, No. 2, pp. 135–152.
6. **Dong, R., O'Mahony, M. P., Schaal, M., McCarthy, K., Smyth, B. (2013).** Sentimental product recommendation. *Proceedings of the 7th ACM Conference on Recommender Systems*, pp. 411–414. DOI: 10.1145/2507157.2507199.
7. **Dong, R., O'Mahony, M. P., Schaal, M., McCarthy, K., Smyth, B. (2016).** Combining similarity and sentiment in opinion mining for product recommendation. *Journal of Intelligent Information Systems*, Vol. 46, No. 2, pp. 285–312. DOI: 10.1007/s10844-015-0379-y.
8. **Dong, R., Schaal, M., O'Mahony, M. P., Smyth, B. (2013).** Topic extraction from online reviews for classification and recommendation. *Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI 13)*, pp. 1310–1316.
9. **Dong, R., Schaal, M., O'Mahony, M. P., McCarthy, K., Smyth, B. (2013).** Opinionated product recommendation. *Case-Based Reasoning Research and Development: 21st International Conference, ICCBR 2013*, Vol. 7969, pp. 44–58. DOI: 10.1007/978-3-642-39056-2_4.
10. **He, R., McAuley, J. (2016).** Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. *Proceedings of the 25th International Conference on World Wide Web*, pp. 507–517. DOI: 10.1145/2872427.2883037.
11. **Hu, M., Liu, B. (2004).** Mining and summarizing customer reviews. *Proceedings of the tenth ACM SIGKDD International Conference on Knowledge discovery and data mining*, pp. 168–177. DOI: 10.1145/1014052.1014073.
12. **Leung, C., Chan, S., Chung, F., Ngai, G. (2011).** A probabilistic rating inference framework for mining user preferences from reviews. *World Wide Web*, Vol. 14, pp. 187–215. DOI: 10.1007/s11280-011-0117-5.
13. **Liu, B. (2012).** Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, No. 1, pp. 1–167. DOI: 10.1007/978-3-031-02145-9.
14. **Liu, B., Hu, M., Cheng, J. (2005).** Opinion observer: Analyzing and comparing opinions on the web. *Proceedings of the 14th International Conference on World Wide Web*, pp. 342–351. DOI: 10.1145/1060745.1060797.
15. **Loper, E., Bird, S. (2002).** Nltk: The natural language toolkit. DOI: 10.3115/1118108.1118117.

16. **Lu, Y., Dong, R., Smyth, B. (2016).** Context-aware sentiment detection from ratings. *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pp. 87–101. DOI: 10.1007/978-3-319-47175-4_6.
17. **Maslowska, E., Malthouse, E. C., Bernritter, S. F. (2017).** The effect of online customer reviews' characteristics on sales. In *Advances in Advertising Research*, pp. 87–100. DOI: 10.1007/978-3-658-15220-8_8.
18. **McAuley, J., Targett, C., Shi, Q., Van Den Hengel, A. (2015).** Image-based recommendations on styles and substitutes. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 43–52. DOI: 10.1145/2766462.2767755.
19. **Medhat, W., Hassan, A., Korashy, H. (2014).** Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, Vol. 5, No. 4, pp. 1093–1113. DOI: 10.1016/j.asej.2014.04.011.
20. **Mejova, Y. (2009).** Sentiment analysis: An overview. Computer Science Department.
21. **Sanner, M. F. (1999).** Python: A programming language for software integration and development. *J Mol Graph Model*, Vol. 17, No. 1, pp. 57–61.
22. **Sharma, R., Bhattacharyya, P. (2013).** Detecting domain dedicated polar words. *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pp. 661–666.
23. **Tallarida, R. J., Murray, R. B. (1987).** Chi-square test. In *Manual of Pharmacologic Calculations*, pp. 140–142. DOI: 10.1007/978-1-4612-4974-0_43.
24. **Taylor, A., Marcus, M., Santorini, B. (2003).** The penn treebank: An overview. *Treebanks: Building and using parsed corpora*, pp. 5–22. DOI: 10.1007/978-94-010-0201-1_1.
25. **Toutanova, K., Klein, D., Manning, C. D., Singer, Y. (2003).** Feature-rich part-of-speech tagging with a cyclic dependency network. *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, Vol. 1, pp. 252–259. DOI: 10.3115/1073445.1073478.
26. **Wallach, H. M. (2004).** Conditional random fields: An introduction. *Technical Reports (CIS)*, pp. 22.
27. **Watson, F., Wu, Y. (2022).** The impact of online reviews on the information flows and outcomes of marketing systems. *Journal of Macromarketing*, Vol. 42, No. 1, pp. 146–164. DOI: 10.1177/02761467211042552.

*Article received on 09/11/2022; accepted on 16/01/2023.
The corresponding author is Pradnya Bhagat.*