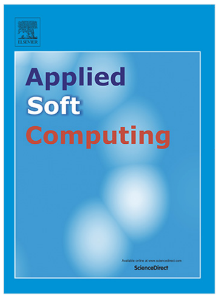
Journal Pre-proof



Applying sentiment analysis to automatically classify consumer comments concerning marketing 4Cs aspects

Hao-Chiang Koong Lin, Tao-Hua Wang, Guo-Chung Lin, Shu-Chen Cheng, Hong-Ren Chen, Yueh-Min Huang

PII: S1568-4946(20)30693-1
DOI: https://doi.org/10.1016/j.asoc.2020.106755
Reference: ASOC 106755

To appear in: *Applied Soft Computing Journal*

Received date : 6 March 2020
Revised date : 4 September 2020
Accepted date : 19 September 2020

Please cite this article as: H.-C.K. Lin, T.-H. Wang, G.-C. Lin et al., Applying sentiment analysis to automatically classify consumer comments concerning marketing 4Cs aspects, *Applied Soft Computing Journal* (2020), doi: https://doi.org/10.1016/j.asoc.2020.106755.

Applying Sentiment Analysis to Automatically Classify Consumer Comments Concerning Marketing 4Cs Aspects

Hao-Chiang Koong Lin[a] ,Tao-Hua Wang[b], Guo-Chung Lin[c], Shu-Chen Cheng[d]*, Hong-Ren Chen[e], Yueh-Min Huang[f]

[a]Department of Information and Learning Technology, National University of Tainan, Taiwan
[b,f]Department of Engineering Science, National Cheng Kung University, Taiwan.
[c]Department of information, Taiwan Sugar Corporation.
[d]Department of Computer Science and Information Engineering, Southern Taiwan University of Science and Technology.
[e]Department of Digital Content and Technology, National Taichung University of Education.

Abstract

With the rapid development of science and technology, consumers are used to searching online for evaluations before purchasing products. Manufacturers can also utilize such information like users' usage habits, browsed websites, comments, messages, etc. to formulate marketing strategies suitable for their products. Several researches developed opinion mining on predicting the polarity of consumers' comments, but few of them were from marketing point of view. In this regards, this study looks to establish an automated way to collect and analyze consumers' comments in social networks, automatically classify them into marketing 4Cs and non-marketing categories from a large number of consumer comments, and divide the category of marketing 4Cs articles into four types of attribute dimensions to analyze emotional polarity. Based on the marketing theory of 4Cs and LDA topic analysis, this study extracted the characteristic keywords from the collected consumer reviews for corpus classification and sentiment polarity analysis. This study further establishes a feature keyword library for specific fields, hoping to improve the accuracy of sentiment analysis through these keywords, simplify the process of consumers' searches for product evaluations, and facilitate consumers to search for helpful target information.

Keyword: Sentiment Analysis, Opinion Mining, LDA, 4Cs of Marketing

1. Introduction

With the widespread usage of social networks, many people participate in the information distribution in the web environment [1][5][19]. Users share opinions about products by providing textual reviews. These opinions emerged as a critical factor of future consumer purchasing decisions [10][12][33]. With the increasing popularity of network and intelligent mobile devices, the mode of network communication has changed from users' passively receiving information transmitted by website operators to users' sending out various information generated by themselves. Such information with emotional opinions has a certain degree of influence on business enterprises, politics, and even individuals [2][28]. Nowadays, consumers are used to web search engines finding product usage evaluations, usage experiences, etc. to help them evaluate whether a product meets their own needs before purchasing it. Therefore, understanding consumer opinions or evaluations through the web is of great help for manufacturers to formulate marketing strategies [19][20], but

* Corresponding author. Tel.: +886-6-2533131.
E-mail address: kittyc@stust.edu.tw.

network information is very complicated and unsystematic. When consumers look for product usage evaluations through keywords, they must click on a large number of web links to open a web page and then read the article thoroughly to identify the evaluation opinions. This is a lengthy process for consumers.

Opinion Mining, which is also named Sentiment Analysis, is a popular technology used to solve this problem in recent years [3][21][41]. Opinion mining is the process of using natural language processing and text analytics to extract information in social networks [29]. Consumer evaluation on the Internet initially exists in the form of text. With the popularity of consumer 3C media, consumers are encouraged to insert photos or other media materials taken by themselves when publishing evaluation articles to form so-called "unpacking articles". The writing style of such evaluation articles is unstructured, and the content of the articles is not just text. In the face of a large number of unstructured articles, it is quite a hard job to capture consumer evaluations by manual processing [4][18]. The development of online comment mining and extraction in the past ten years is one of new research fields [20][36]. The obtained evaluation articles can be processed in an automated way through opinion mining or sentiment analysis technology, thereby analyzing the tendency of evaluation opinions [16][17][32].

Liu [6] mentioned in sentiment analysis or opinion mining research that three levels of sentiment analysis can be carried out for consumer comment articles: document level, sentence level, and aspect level. Most sentiment analysis research focuses on the overall content of the text - that is, the emotional scores of all sentences in the whole article are analyzed at the document level and sentence level [24]. However, general consumers are concerned about certain aspects of evaluations in product review articles, such as price, specification, appearance, etc [25]. Therefore, this research targets to establish an automated way to collect and analyze online consumer reviews. After the reviews are processed through sentiment analysis, it is convenient for consumers to quickly obtain the evaluation of a product in different aspects when searching for a target product. Consumers do not need to spend a lot of time reading the entire content before they know the positive and negative evaluations of the product in such an article.

For product evaluations to be searched, the purpose of this study is to automatically classify review articles into marketing 4Cs and non-marketing from a large number of consumer reviews and then to analyze emotional polarity according to different attribute aspects of marketing 4Cs, of which the attribute aspects are divided into 4 categories according to the marketing theory of 4Cs. In addition, this study establishes a feature keyword library belonging to a specific field, hoping to improve the process of consumers searching for product evaluations through these experiments, enhance the accuracy of sentiment analysis, and facilitate consumers to search for helpful target information. Moreover, the topics discussed by netizens will differ slightly in different time ranges, so the corpus collected will also produce diverse topics. To develop a semi-automated approach to dig out the popular comment topics of consumers, this study proposed to analyze through LDA of topic model before sentiment analysis. It can greatly reduce labor and time costs to expand and build a characteristic word bank.

In the rest sections of this paper, section 2 introduces related literature reviews of this study. Section 3 describes system architecture of this study. In section 4 is the methods of this study, including data collection, manual classification, data pre-processing, topic analysis, corpus classification, and sentiment analysis used in this study. Section 5 analyses the experimental results of research method used in this study. Section 6 summarize conclusion of this study.

2. Literature Review

Sentiment analysis uses Natural Language Processing (NLP) [16][42], Information Retrieval (IR), and structured and unstructured data mining to analyze subjective information in articles [7]. Isidoros and Ioannis [8] pointed out that the purpose of sentiment analysis is to understand the polarity of opinions expressed by people on a topic through different media. Sentiment analysis is a concept that covers many tasks [26], such as extracting emotions, emotion classification, subjective classification, opinion summary or opinion spam detection, etc. [9][23] and is used to determine whether the emotions expressed in the text are positive or negative [35][38]. The mass is relying on online and the significance of sentiment analysis in daily lives is increasing [43].

Sentiment analysis mainly obtains emotion polarity by identifying the matching relationship between characteristic words and emotion words in the text [21][22][25][30] and relies in detecting the hidden subjective expression in the text [40]. In [39], the authors indicated that automated analysis has poorer performance than human raters on more difficult and noisy data sets. Emadi and Rahgozar found that pure NLP-based methods have very low accuracy and these methods need to be fused with machine learning–based methods to achieve higher accuracy [44]. Kim, Gupta and Rho used the LDA algorithm to create a cluster of scientific themes to generate issue keywords and the precisions of the top-5 retrieved terms with search query are about 75% [1]. This study uses the unsupervised learning method of Latent Dirichlet Allocation (LDA) as shown in Figure 1.

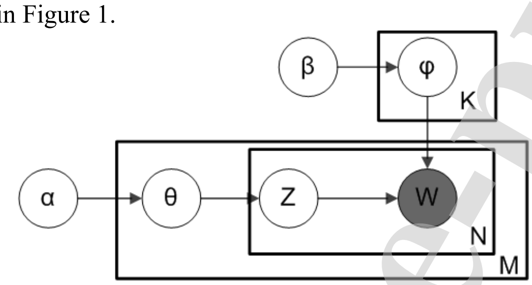


Figure 1. LDA Bayesian Network Structure [37]

In the LDA model the word generation method for the theme of the document is as shown in the above figure structure [27][31]. In Figure 1, M is the total number of all documents, n is the total number of all words, k is the total number of all topics, α is the distribution of the theme in the document, β is the distribution of the words in the theme, θi is the distribution of the theme in document i, φk is the distribution of the words in theme k, Zij is the distribution of the j^{th} word in the theme in the i^{th} document, and Wij indicates that word j in document i is the finally generated word.

LDA topic models are usually implemented through two methods: Variantional Methods or Gibbs Sampling [23]. Because Gibbs Sampling is easier to perform LDA topic analysis and expand, it can get a better approximation quickly [11], and thus this study decided to use the Gibbs Sampling method to carry out LDA topic analysis. The marketing theory of 4Cs was put forward by the U.S. marketing expert Dr. Robert F. Lauterborn in 1990 [15]. It takes consumer demand as the orientation and resets four basic elements of the marketing mix: consumer, cost, convenience, and communication. Based on the marketing theory of 4Cs, this study has established a semi-automated feature lexicon of different product aspects, which is matched with the National Taiwan University Sentiment Dictionary (NTUSD). In this way, a search information process suitable for consumers to quickly grasp positive and negative reviews of products can be established, and the cost of manual participation can be reduced.

3. System Architecture

The research structure is mainly divided into topic analysis, corpus classification, and sentiment analysis. Chinese corpus was subject to topic analysis after pre-processing, and the topic analysis was used to establish the characteristic keyword database for each aspect of marketing 4Cs. The corpus classification and sentiment analysis were then carried out for consumer reviews through the lexicon established in this research. The overall system architecture is shown in Figure 2. The steps to process and analyse the applied data-set are data pre-processing, topic analysis, corpus classification, and sentiment analysis. The detail descriptions are illustrated in section 4.3~4.6.

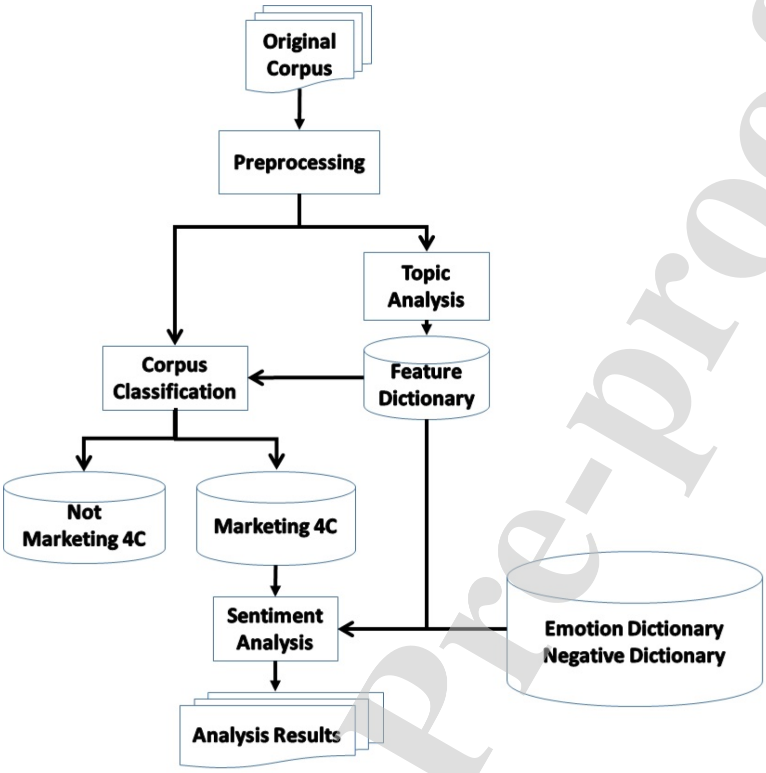


Figure 2. System Architecture and Flowchart

4. Method

4.1. Data Collection

The sources of car reviews are very diverse in Taiwan. NTU PTT is the largest electronic bulletin board in Taiwan. The forums are subdivided into boards. In order to focus on the source attributes of consumers, this study chose the Car Board of NTU PTT as the data source. This study used R language to write a web crawler and collected 2,837 reviews from the PTT Car Board in a month for topic analysis and sentiment analysis. The data format of a post from the PTT Car Board contains the information of the post date, the author's name, the title and the text of the post, approval or disapproval about the post, etc. Texts of the posts and the information of "approval or disapproval about the post" were used in this study for topic analysis and sentiment analysis. The detail procedures of data preprocessing are described in Section 4.3. In addition, the dictionary method was used for sentiment analysis, and emotional lexicon and negative lexicon were needed to mine emotion words and negative words in Chinese corpus. Thus, NTUSD was collected as the emotional lexicon for this study, and a negative lexicon was established from the negative lexicon provided by relevant studies.

4.2. Manual Classification

In order to obtain consumers' comments, this study adopted the marketing theory of 4Cs and sorted out the aspects that consumers care about through the four major aspects of marketing 4Cs. According to the marketing theory of 4Cs and collecting the contextual attributes of data sources, this study sorted out the topics included in each aspect of marketing 4Cs as follows.

Table 1 Issues of Various Aspects of Marketing 4Cs

Aspect	Topic direction
Consumer	Consumer demand issues, focusing on the nature of the product such as appearance, vehicle type, general equipment, safety equipment, interior decoration, etc.
Cost	Price issues, such as CP value, cost performance, price, insurance, contract payment, etc.
Convenience	Issues related to convenience, service, or access.
Communication	Communication issues, such as advertising, publicity and marketing, brand, promotion, etc.

In order to check the accuracy rate of the research results, the collected articles were manually classified and marked for the subsequent accuracy rate test. Consumer review articles were distributed to three people for marking through random sampling for manual classification. When the sentences met the 4Cs topic direction, they were judged as 4Cs marketing category. In addition to being classified as 4Cs marketing category, the three people respectively marked the emotional polarity of the articles as four categories of "positive", "negative”, “neutral", and "no emotion”. The results of the three-person classification were determined by a majority vote. If the classification results were different, then the three people discussed to determine to which category they belonged.

4.3 Data Preprocessing

Before topic analysis and sentiment analysis of Chinese corpus, unnecessary corpus was removed, and the data were structured for subsequent emotional calculation and analysis. The "author" and "title" fields in the text data captured from the PTT Car Board by using web crawler can effectively help subsequent analysis. As for other normal words that have no meaning for the subsequent analysis, they were cleared at this stage. At the same time, in order to improve the efficiency of subsequent analysis, English letters were changed to lower-case.

Because Chinese words are not like English words, there is no space or other symbol separation between words, and the part of speech must be determined in a special way and then separated. In this study, the jiebaR package of R language was used to process Chinese word segmentation, which is highly convenient and extensible.

There are many words created by netizens in the articles of the PTT Car Board; for example, "4 circles" refers to the automobile brand Audi. In addition, before Chinese word segmentation processing, wrong words in the Chinese corpus must be corrected, and these words should be extended to the custom word library of the jiebaR package, so that the system can correctly judge the words to be segmented and thus avoid information loss caused by these wrong words in the topic analysis stage.

Negative words were excluded from the pre-set stop word bank of the jiebaR package to prevent some negative words from being lost due to word segmentation operations. At the same time, the top 100 words with the highest frequency of "balanced corpus word set and word frequency statistics of Academia Sinica" and the special words of PTT netizens were added, while words with less correlation in this study were expanded to the stop word bank of jiebaR package in the pre-processing stage.

4.4 Topic Analysis

At this stage, LDA topic analysis was carried out to produce candidate words that can be marked as marketing 4C-related characteristic keywords. However, the research object was articles discussed in the PTT Car Board. These topic words are not necessarily expressed in terms of nouns, but may also be other parts of speech. Therefore, a part of speech was not limited in order to analyze the topics discussed more completely.

Relevant parameters in LDA analysis include: number of Topics, Topic Words, and Iterations, of which Topic number has the greatest influence. To set the number of Topics, this study used the changes of perplexity and loglikelihood for the estimate. Perplexity decreases with the increase of Topic number, while loglikelihood increases with the increase of Topic number. Generally speaking, the number of Topics when the two changes tend to be flat can be used as the estimated number of Topics. If only one modeling is performed, then the perplexity and loglikelihood values cannot verify the validity of the model, and so a 10-fold cross-validation test was adopted. The LDA experiment used the Gibbs sampling mode to find the best value of Topic, and the number of Topic Words was also set to 10 - that is, 10 words were generated as candidate words for each Topic.

In order to discuss the aspect issues that consumers care about, this study is based on the marketing theory of 4Cs to set the multi-aspects of products that consumers care about. Topic and its Topic Words were found as candidate words, which were manually marked to generate the characteristic words of the four major aspects of marketing 4C, so as to facilitate the subsequent comparison and examination of the correct rate of use. The characteristic keywords defined in this study, i.e. attribute words or characteristic words in related research studies of sentiment analysis, were used as the target of the author's emotion opinions in the comment corpus. Candidates generated from LDA topic analysis results were marked by three people to judge which aspect of marketing 4Cs these candidates belong to respectively. If they did not meet any of the four aspects of marketing 4Cs, then they were marked as "other". The marking results of the three people were decided by a majority vote. If the markings of the three people were different from each other, a fourth person was invited to join the discussion and determine the jointly approved marking results.

4.5 Corpus Classification

This study used the word bank comparison method to classify the corpus and classified the corpus articles according to the marketing 4Cs characteristic keyword database. As long as the sentences in the corpus articles conformed to the words in the characteristic keyword database, they can be classified as this aspect and an article might be sorted to multiple marketing aspects. The classification results of this study helped establish a Confusion Matrix, and Accuracy, Precision, Recall, and F1-Measure were used as the measurement of the classification effect.

4.6 Sentiment Analysis

At this stage, the sentiment analysis lexicon was introduced and compared with the characteristic keywords generated by topic analysis to find out each emotional word and negative word, and then the emotional score was calculated according to the set formula. According to scholars, even if the same emotional word has the opposite emotional polarity in different fields [34], it is mainly because the emotional word may change the emotional polarity according to the context [13]. In addition, NTUSD needs to be additionally expanded, because it lacks emotional vocabulary in the automobile field.

In this study, the characteristic keywords and emotion words marked as marketing 4Cs were used to calculate the distance of matching, and the positive and negative polarities of the emotion dictionary were combined to confirm whether the comment sentence gave the emotion score weight. This study found that the matched emotion words are better when the position distance between characteristic keywords and emotion words is roughly within 5-unit words. If negative words are found within a certain range of words, then the polarity of the emotion words can be changed [14]. This study also found that negative words are roughly within the distance between 2-unit words of emotion words. Therefore, according to the above matching distance, this study proposed the following formula to calculate the emotion scores of corpus sentences:

Emotion score = polarity weight of emotion words × negative word weight

(a). If there is a matching relationship between the characteristic keywords and emotion words in the sentence of the corpus, then the score is given according to the polarity of emotion words: 1 for positive emotion words and -1 for negative emotion words.

(b). If the negative words and emotion words in the corpus sentence reach a matching relationship, it indicates that the emotional polarity of the sentence is opposite, and so the negative word weight score is -1, and the opposite polarity score can be obtained by multiplying the emotion word score.

After all sentences were analyzed according to the aspects of the characteristic keyword, the emotion score of each aspect of the sentence can be obtained. The emotional scores of the four major aspects of the same article for all sentences respectively were added up to obtain the emotion scores of the four major aspects of the article marketing 4C. According to the emotion scores after the adding, the emotional polarity can be divided into four categories as follows.

(a). If the total score is greater than 0, it is "positive emotion".
(b). If the total score is less than 0, it is “negative emotion”.
(c). If the emotion score of a sentence is not equal to 0, but the total score is equal to 0, then the aspect is "neutral emotion".
(d). If no emotion score appears in any sentence, then the aspect is "no emotion".

5. Experimental Results

5.1. Manual Marking Results

Figure 3 is the histogram of manual marking results for 4Cs articles in marketing. It is evident that the articles classified under the aspects of Consumer and Convenience were over 80% in the test samples, and the articles classified under the aspect of Cost were over 50%. It is obvious that the proportion was the lowest for the articles classified under the aspect of Communication. It can be seen that most users of the PTT Car Board were more enthusiastic about discussing consumer demand for auto products and related maintenance or services. As for price-related issues, their degree of discussion ranked third, while advertising or sales-related issues were the lowest.

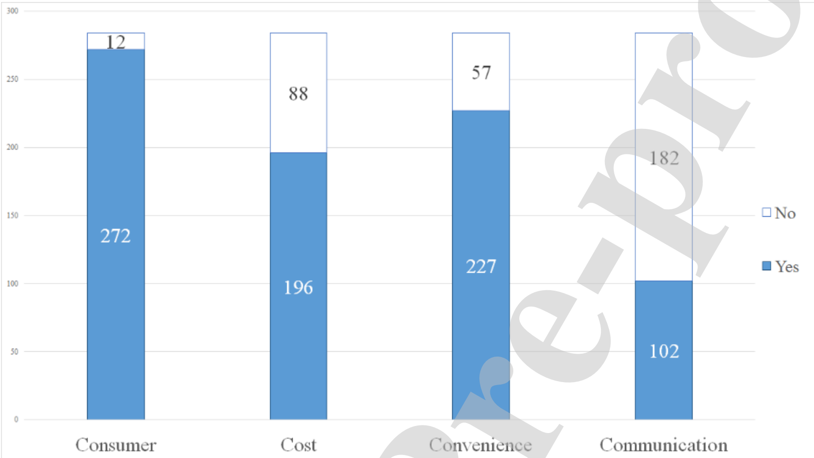


Figure 3. Manual Marking Results of Marketing 4Cs Articles

Figure 4 is a histogram of emotion polarity manually marked from the articles classified as marketing 4Cs. It can be seen that compared with positive and negative emotions in the four major aspects, negative emotion articles are more than positive emotion articles. The inference should be that netizens often publish negative emotion articles in the PTT Car Board. In particular, many aggressive people are nicknamed "keyboard drivers" on the Car Board and often criticize with the contents of irony semantics, thus gradually forming an emotional tendency for PTT netizens to publish articles.

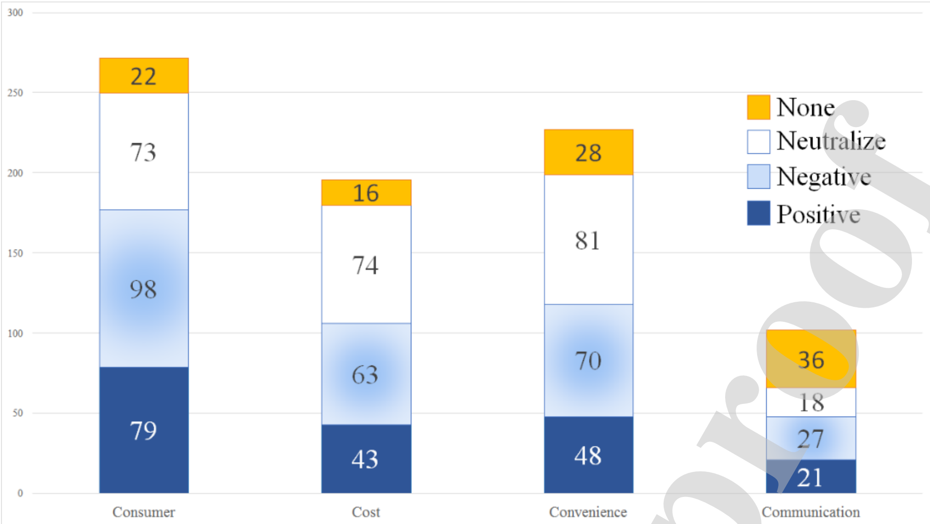


Figure 4. Manual Marking Results of Marketing 4Cs Articles

5.2. Results of Topic Analysis

The LDA experimental data source is pre-processed consumer review articles. TF (Term Frequency) was taken as the word frequency weight and used as the dataset analyzed by the LDA algorithm. The Gibbs sampling mode was adopted for the algorithms, and each Topic generated 10 words as candidate words. In order to meet statistical reliability, it is necessary to compare the change results of model perplexity and log likelihood value averaged under different Topic numbers. Therefore, this study used the 10-fold crosscheck method to estimate the results, as shown in Figure 5. The 10-fold crosscheck of this system is executed in a circular manner three times - namely, fold 1, fold 2, and fold 3. It can be seen from the figure that when Topic number = 50, the perplexity number is the smallest, indicating that the most suitable topic number for LDA topic analysis is 50 for corpus articles collected in this study.

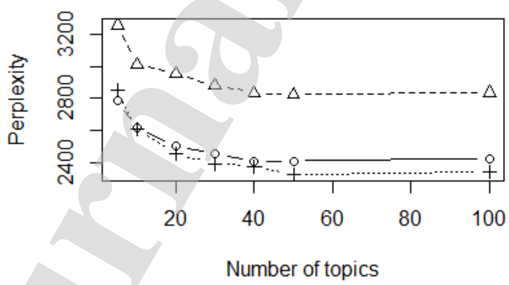


Figure 5. 10-fold Crosscheck Results

If it did not conform to the four major aspects of marketing 4C, then it was marked as "other". The marking results of the three people was decided by a majority vote. If the marking results of the three people were all different, then the fourth person was invited to join in the discussion to determine the commonly accepted marking results. The three people then marked to which aspect of marketing 4Cs the candidate words belonged. After the marking was completed, in order to confirm whether the marking results of the three people were consistent or not, SPSS software was used to carry out Kappa verification in this study. The results of the verification for the marking result consistency are shown in table 2. Kappa verification results were all above 0.9, which shows that the marking results of the three reporters were highly consistent with each other. Although there were a few cases not completely the same, they accounted for a very small proportion of the total. Therefore, the aspects marking of LDA candidate topic words is trustworthy.

Table 2 Kappa Test Results of Candidate Topic Words

Marker	Results of the Verification
A*B	.919
A*C	.947
B*C	.917

5.3 Corpus Classification and Experimental Results

After the candidate words generated in the topic analysis stage were manually marked, the characteristic keywords of the marketing 4Cs face were obtained. The characteristic keywords were then used to judge the aspect to which the article belonged, and the lexicon comparison method was adopted as the corpus classification method.

From the results of Table 3, it can be found that when the four major aspects of this experiment were classified by the system, F1-Measure had an average of 93.08%, which indicated that the classification performance had good effect. The Precision values of the four major aspects were classified by the thesaurus comparison method in this experiment, and the characteristic keywords in this table were obtained through sentiment analysis and manual labeling in a semi-automated manner. As long as there were characteristic words that conformed to this aspect in the article, the system can classify them as this aspect, and so the Precision values were all 100%.

The Precision value of the aspect of Communication was only 88.03%, and the recall value was only 66.67%, which had a gap compared with the results of the other three aspects. Analyzing the reasons, this study defined the issues of Communication aspect as including communication between manufacturers and consumers, advertising promotion, or sales, while there were few topics on the PTT Car Board to discuss the Communication aspect, which resulted in only 7 characteristic keywords of the Communication aspect produced in the topic analysis experiment of this research. The Precision value and Recall value were not high enough, because there were not enough comparable pairs of characteristic keywords in the system classification. If the number of characteristic words can be increased by raising the Communication aspect, then the classification accuracy of Communication orientation should be improved.

This study is using the confusion matrix to evaluate the accuracy, precision, recall, and F1-measure. The four measures of the confusion matrix are described as follows:

- True Positives (TP): the number of true positives.
- False Positives (FP): the number of false positives.

- True Negatives (TN): the number of true negatives.
- False Negatives (FN): the number of false negatives.

where formula of accuracy, precision, recall, and F1-measure are as follows:

$$\text{Accuracy} = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{1}$$

$$\text{Precision} = \frac{TP}{(TP + FP)} \tag{2}$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \tag{3}$$

$$\text{F1} - \text{Measure} = \frac{2 \times Precision \times Recall}{(Precision + Recall)} \tag{4}$$

Table 3 Accuracy Rate of Marketing 4Cs Classification by the System

	Accuracy	Precision	Recall	F1-Measure
Consumer aspect	93.66%	100%	93.38%	96.58%
Cost aspect	97.89%	100%	96.94%	98.45%
Convenience aspect	95.77%	100%	94.71%	97.28%
Communication aspect	88.03%	100%	66.67%	80%
Average	93.84%	100%	87.93%	93.08%

5.4 Sentiment Analysis

Because NTUSD lacks emotion words commonly used by PTT Car Board users, this study employed manual retrieval to find emotion words commonly used by Car Board users from the collected consumer corpus articles and then expanded them into the emotional lexicon of this study. Because some words have ironic meanings, this study adjusted the emotional polarity of some emotion words in NTUSD. The expanded emotional lexicon contains 2850 positive words and 8337 negative words, totaling 11,187 words.

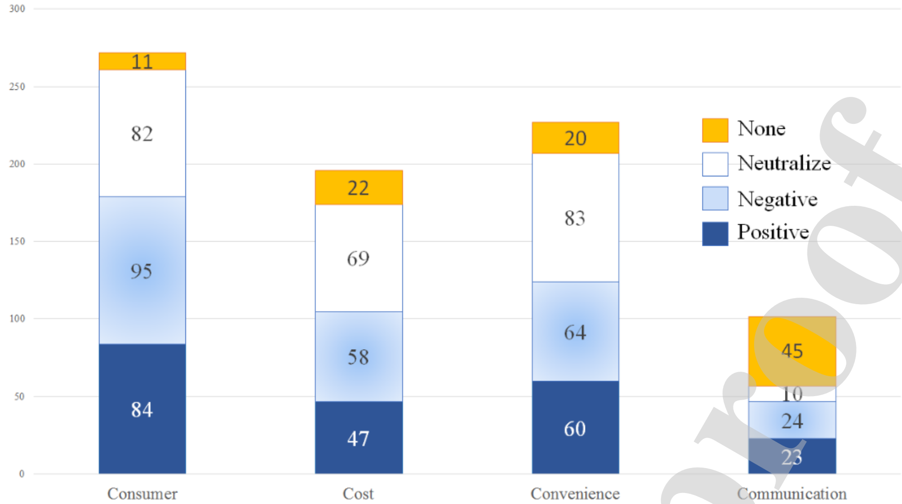


Figure 6. Results of the Automatic Classification of Emotion Polarity

From the results shown in Figure 6, the articles with emotional polarity of the Consumer aspect, i.e. positive, negative, and neutral emotions, accounted for 96% of the test corpus. It can be seen that PTT car users mostly discussed their experiences or consumer needs on the advantages and disadvantages of product use when publishing articles and more often expressed their personal emotions. The comments on articles for the aspects of Cost and Convenience with emotional polarity (positive, negative, and neutral emotions) also reached nearly 90%, indicating that netizens also had a considerable degree of enthusiasm for discussing the cost-effectiveness of automobiles and the convenience of product use or consumption. As for the aspect of Communication, only 57% of comments with emotional polarity in the test corpus show that netizens did not disclose personal emotions when discussing the publicity, advertising, or marketing of automobiles.

Table 4 Accuracy Rate of the System Classification of Emotional Polarity

	Accuracy	Precision	Recall	F1-Measure
Consumer aspect	93.75%	94.27%	98.80%	96.48%
Cost aspect	92.85%	97.70%	94.44%	96.04%
Convenience aspect	90.31%	92.75%	96.48%	94.58%
Communication aspect	87.25%	96.49%	83.33%	89.43%
Average	91.04%	95.30%	93.26%	94.13%

Judging from the accuracy rate of Table 4 sentiment classification, the sentiment analysis experiment in this research is aimed at the effectiveness of marketing 4Cs articles in judging whether they have sentiment polarity. F1-Measure for the four aspects had an average of 94.13%, which shows that the classification effect achieved good performance. The following reasons are observed for articles with inconsistent manual marking and automatic classification in the test data.

(a) Interrogative Semantic Articles

Some of the articles are those where netizens ask questions about 4Cs marketing problems. They are semantically interrogative and do not have emotional polarity. On the other hand, some articles are written unstructured and do not necessarily have punctuation marks, but semantically they can be seen as interrogative and are classified as "no emotion" during manual classification. However, the system classification was less accurate, because punctuation marks were removed during pre-processing of articles and emotional values were obtained as long as the contents conformed to the sentiment analysis formula.

(b) Purely Descriptive Articles

Some articles are in the field of automobiles, but their content is purely to describe certain events or topics, such as news reprinted by netizens about the reckless people on the road (meaning the unruly female, elderly, and older female pedestrians on the road), news unrelated to automobile topics, etc., or teaching articles about certain products or automobiles. The sentences of these articles were classified as having emotional polarity in the system classification, because they just contained characteristic keywords and emotion words matching the matching distance. However, when they were manually marked, they can be considered as news or teaching articles, which are semantically inconsistent with the positive and negative emotions of this study, and so they are judged as having no emotion.

(c) Not Enough Characteristic Keywords

Since the characteristic keywords of this study were analyzed through LDA themes, some words may not be extracted due to word frequency or other factors, resulting in the system classification judgment of no emotion. However, when manually marking, the words are regarded as characteristic words and given positive and negative emotion classifications, resulting in inconsistent results between the two.

(d) Irony

Some articles have ironic meanings, but there was no accurate judgment mechanism in this research experiment, which led to inconsistent results between automatic classification and manual classification.

(e) Different Opinions on Comments

The same review article may compare the advantages and disadvantages of different products in the same aspect. At that time, when consumers view the article from different angles, the article will have different emotional polarities.

6 Conclusion

As shown from the literature review, the way of establishing characteristic words in different fields in sentiment analysis research was mostly done by manually selecting from a large number of corpus articles according to the research topics set in advance. In addition, the topics discussed by netizens will differ slightly in different time ranges, and the corpus collected from analysis will also produce different topics. Therefore, this study proposed to analyze through LDA of topic model before sentiment analysis, so as to dig out the popular comment topics of consumers in a certain period and further used the characteristic keywords analyzed by LDA to classify the corpus through lexicon comparison. From the experimental results, the accuracy of classifying the corpus of consumer comments into 4Cs categories of marketing was 93.84% on average with F1-Measure measurement results of 93.08%. This implies that the system classification of consumer reviews in marketing or non-marketing 4Cs articles in this experiment is good. Moreover, it can greatly reduce labor and time costs by replacing manual marking with a semi-automated approach to expand and build a characteristic word bank.

In order to find out whether the classification of emotional polarity in different aspects of product attributes has a certain degree of accuracy, this study used Lexicon-based methods to conduct a sentiment analysis experiment. From the experimental results, it can be observed that the accuracy of whether consumer reviews have emotional classification in different aspects was 91.04% on average, and the F1-Measure measurement effect was 94.13% on average. The accuracy of positive and negative sentiment classification was 91.74% on average, and the F1-Measure measurement effect was 90.68% on average. This proved that the experimental system had good accuracy in classifying sentiment polarity and provides consumers with reliable sentiment analysis results of the product attributes for various aspects after consumers search out 4Cs types of marketing review articles, so that consumers do not need to spend too much time browsing the contents of articles to analyze product usage experiences.

According to the above research conclusion, the contribution of this research can be summarized in the following 2 aspects.

(a). Establishing Product Attribute Multi-aspect Characteristic Words in A Semi-automated Way

Based on the marketing theory of 4Cs and LDA topic analysis, this study is able to extract the most representative themes and characteristic keywords from the collected consumer reviews during this period, which can be used as the characteristic basis for corpus classification and sentiment analysis, so that the analysis results are more in line with the product usage views that consumers care about.

(b). Empirical Analysis of Sentiment Analysis Combined with Attribute Multi-aspect Classification Model

Through the comparison between manual classification and automatic classification, it is proved that the corpus classification and sentiment polarity classification of product attributes in this study exhibit reliable accuracy. The manual classification replaced by the topic model combined with sentiment analysis can provide a more convenient product review search and enable consumers to quickly obtain and analyze search results.

This study established a feature keyword library for specific fields. The research limitation is that a certain feature keyword library is needed for specific fields and there is no general keyword library for all fields. Besides, if the consumer’s review contains interrogative or ironic meaning, it is hard to accurately classify the review to the correct emotional polarity. In the future work, the semantic analysis for sarcasm detection, rumor detection, irony detection can be implemented to enhance the accuracy of sentiment analysis.

References

[1] M. Kim, B. B. Gupta, S. Rho, Crowdsourcing based scientific issue tracking with topic analysis, Applied Soft Computing. 66 (2018) 506-511.

[2] D. M. E. D. M. Hussein, A survey on sentiment analysis challenges, Journal of King Saud University-Engineering Sciences. 30 (4) (2016) 330-338.

[3] W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: A survey, Ain Shams engineering journal. 5 (4) (2014) 1093-1113.

[4] K. Khan, B. Baharudin, A. Khan, A. Ullah, Mining opinion components from unstructured reviews: A review, Journal of King Saud University-Computer and Information Sciences. 26 (3) (2014) 258-275.

[5] F. Ali, K. S. Kwak, Y. G. Kim, Opinion mining based on fuzzy domain ontology and Support Vector Machine: A proposal to automate online review classification, Applied Soft Computing. 47 (2016) 235-250.

[6] B. Liu, Sentiment analysis and opinion mining, Synthesis lectures on human language technologies. 5 (1) (2012) 1-167.

[7] K. Ravi, V. Ravi, A survey on opinion mining and sentiment analysis: tasks, approaches and applications, Knowledge-Based Systems. 89 (2015) 14-46.

[8] Perikos, I. Hatzilygeroudis, Recognizing emotions in text using ensemble of classifiers, Engineering Applications of Artificial Intelligence. 51 (2016) 191-201.

[9] J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, E. Herrera-Viedma, Sentiment analysis: A review and comparative analysis of web services, Information Sciences. 311 (2015) 18-38.

[10] G. Cosma, G. Acampora, A computational intelligence approach to efficiently predicting review ratings in e-commerce, Applied Soft Computing. 44 (2016) 153-162.

[11] G. Heinrich, Parameter estimation for text analysis, Technical report. (2009).
[12] C. Catal, M. Nangir, A sentiment classification model based on multiple classifiers, Applied Soft Computing. 50 (2017) 135-141.
[13] H. Kansal, D. Toshniwal, Aspect based summarization of context dependent opinion words, Procedia Computer Science. 35 (2014) 166-175.
[14] J. H. Wang, C. C. Lee, Unsupervised opinion phrase extraction and rating in Chinese blog posts, In 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, 2011, pp. 820-823.
[15] T. Paul, P. Bihani, Expectation Based Customer Oriented Marketing Mix- A Conceptual Framework. 3 (1) (2014) 51-54.
[16] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, Sentiment analysis of twitter data, In Proceedings of the Workshop on Language in Social Media, 2011, pp. 30-38.
[17] Pang, L. Lee, S. Vaithyanathan, Thumbs up?: sentiment classification using machine learning techniques, In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, 2002, pp. 79-86.
[18] E. Haddi, X. Liu, Y. Shi, The role of text pre-processing in sentiment analysis, Procedia Computer Science. 17 (2013) 26-32.
[19] J. Jin, P. Ji, C. K. Kwong, What makes consumers unsatisfied with your products: Review analysis at a fine-grained level, Engineering applications of artificial intelligence. 47 (2016) 38-48.
[20] N. Naveed, T. Gottron, J. Kunegis, A. C. Alhadi, Bad news travel fast: A content-based analysis of interestingness on twitter, In Proceedings of the 3rd international web science conference, 2011, pp. 8.
[21] A. Sharma, S. Dey, A comparative study of feature selection and machine learning techniques for sentiment analysis, In Proceedings of the 2012 ACM research in applied computation symposium, 2012, pp. 1-7.
[22] T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis, Computational linguistics. 35 (3) (2009) 399-433.
[23] C. Lin, Y. He, R. Everson, S. Ruger, Weakly supervised joint sentiment-topic detection from text, IEEE Transactions on Knowledge and Data engineering. 24 (6) (2011) 1134-1145.
[24] Y. Fang, H. Tan, J. Zhang, Multi-strategy sentiment analysis of consumer reviews based on semantic fuzziness, IEEE Access. 6 (2018) 20625-20631.
[25] L. Lizhen, S. Wei, W. Hanshi, L. Chuchu, L. Jingli, A novel feature-based method for sentiment analysis of Chinese product reviews, China communications. 11 (3) (2014) 154-164.
[26] R. Xia, F. Xu, C. Zong, Q. Li, Y. Qi, T. Li, Dual sentiment analysis: Considering two sides of one review, IEEE transactions on knowledge and data engineering. 27 (8) (2015) 2120-2133.
[27] Y. Li, X. Zhou, Y. Sun, H. Zhang, Design and implementation of Weibo sentiment analysis based on LDA and dependency parsing, China Communications. 13 (11) (2016) 91-105.
[28] C. H. Tai, Z. H. Tan, Y. S. Chang, Systematical approach for detecting the intention and intensity of feelings on social network, IEEE journal of biomedical and health informatics. 20 (4) (2016) 987-995.
[29] R. Xia, J. Jiang, H. He, Distantly supervised lifelong learning for large-scale social media sentiment analysis, IEEE transactions on affective computing. 8 (4) (2017) 480-491.
[30] L. C. Yu, J. Wang, K. R. Lai, X. Zhang, Refining word embeddings using intensity scores for sentiment analysis, IEEE/ACM Transactions on Audio, Speech, and Language Processing. 26 (3) (2017) 671-681.
[31] J. D. Zhang, C. Y. Chow, Crats: An lda-based model for jointly mining latent communities, regions, activities, topics, and sentiments from geosocial network data, IEEE Transactions on Knowledge and Data Engineering. 28 (11) (2016) 2895-2909.
[32] A. Valdivia, M. V. Luzón, F. Herrera, Sentiment analysis in tripadvisor. IEEE Intelligent Systems, 32 (4) (2017) 72-77.
[33] V. García-Díaz, J. P. Espada, R. G. Crespo, B. C. P. G-Bustelo, J. M. C. Lovelle, An approach to improve the accuracy of probabilistic classifiers for decision support systems in sentiment analysis, Applied Soft Computing. 67 (2018) 822-833.
[34] A. Moreno-Ortiz, J. Fernández-Cruz, Identifying polarity in financial texts for sentiment analysis: a corpus-based approach, Procedia-Social and Behavioral Sciences. 198 (2015) 330-338.
[35] M. Fernández-Gavilanes, T. Álvarez-López, J. Juncal-Martínez, E. Costa-Montenegro, F. J. González-Castaño, Unsupervised method for sentiment analysis in online texts, Expert Systems with Applications. 58 (2016) 57-75.
[36] C. Bucur, Using opinion mining techniques in tourism, Procedia economics and finance. 23 (2015) 1666-1673.
[37] J.R. Saura, D. Bennet, A Three-Stage Methodological Process of Data Text Mining: A UGC Business Intelligence Analysis, Symmetry-Basel. (2019) 1-21. doi: 10.13140/RG.2.2.11093.06880.
[38] Al-Moslmi, T., Albared, M., Al-Shabi, A., Omar, N., & Abdullah, S. (2018). Arabic senti-lexicon: Constructing publicly available language resources for Arabic sentiment analysis. Journal of information science, 44(3), 345-362.
[39] Kirilenko, A. P., Stepchenkova, S. O., Kim, H., & Li, X. (2018). Automated sentiment analysis in tourism: Comparison of approaches. Journal of Travel Research, 57(8), 1012-1025.
[40] Ragini, J. R., Anand, P. R., & Bhaskar, V. (2018). Big data analytics for disaster response and recovery through sentiment analysis. International Journal of Information Management, 42, 13-24.
[41] Li, W., Jin, B., & Quan, Y. (2020). Review of Research on Text Sentiment Analysis Based on Deep Learning. Open Access Library Journal, 7(3), 1-8.
[42] Boudad, N., Faizi, R., Thami, R. O. H., & Chiheb, R. (2018). Sentiment analysis in Arabic: A review of the literature. Ain Shams Engineering Journal, 9(4), 2479-2490.
[43] Kumar, A., & Jaiswal, A. (2020). Systematic literature review of sentiment analysis on Twitter using soft computing techniques. Concurrency and Computation: Practice and Experience, 32(1), e5107.

[44] Emadi, M., & Rahgozar, M. (2020). Twitter sentiment analysis using fuzzy integral classifier fusion. Journal of Information Science, 46(2), 226-242.

Credit Author Statement

Hao-Chiang Koong Lin: Conceptualization, Methodology, Formal analysis
Tao-Hua Wang: Writing- Original draft preparation, Investigation, Visualization
Guo-Chung Lin: Writing- Original draft preparation, Software
Shu-Chen Cheng: Software, Writing- Reviewing and Editing, Formal analysis
Hong-Ren Chen: Conceptualization, Methodology
Yueh-Min Huang: Conceptualization, Supervision

Highlights

1. Consumers share opinions and search online for evaluations before purchasing products.
2. Few researches analyze the polarity of consumers’ comments from marketing point of view.
3. This study analyzed consumers’ comments and automatically classify them into marketing 4Cs and non-marketing categories.
4. This study analyzed emotional polarity for the marketing 4Cs articles.
5. This study established a feature keyword library for specific fields.
6. This study simplified the consumers’ searches for product evaluations and helpful target information.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: