Journal of Information & Knowledge Management
Vol. 18, No. 3 (2019) 1950033 (18 pages)

DOI: 10.1142/S0219649219500333

Effectiveness of Domain-Based Lexicons vis-à-vis General Lexicon for Aspect-Level Sentiment Analysis: A Comparative Analysis

Madan Lal Yadav* and Basav Roychoudhury†
Indian Institute of Management Shillong
Nongthymmai, Shillong 793014, Meghalaya, India
**madan.fpm14@iimshillong.ac.in*
†brc@iimshillong.ac.in

Published 13 August 2019

Abstract. One can either use machine learning techniques or lexicons to undertake sentiment analysis. Machine learning techniques include text classification algorithms like SVM, naive Bayes, decision tree or logistic regression, whereas lexicon-based sentiment analysis uses either general or domain-based lexicons. In this paper, we investigate the effectiveness of domain lexicons vis-à-vis general lexicon, wherein we have performed aspect-level sentiment analysis on data from three different domains, viz. car, guitar and book. While it is intuitive that domain lexicons will always perform better than general lexicons, the actual performance however may depend on the richness of the concerned domain lexicon as well as the text analysed. We used the general lexicon SentiWordNet and the corresponding domain lexicons in the aforesaid domains to compare their relative performances. The results indicate that domain lexicon used along with general lexicon performs better as compared to general lexicon or domain lexicon, when used alone. They also suggest that the performance of domain lexicons depends on the text content; and also on whether the language involves technical or non-technical words in the concerned domain. This paper makes a case for development of domain lexicons across various domains for improved performance, while gathering that they might not always perform better. It further highlights that the importance of general lexicons cannot be underestimated — the best results for aspect-level sentiment analysis are obtained, as per this paper, when both the domain and general lexicons are used side by side.

Keywords: Domain lexicon; SentiWordNet; aspect-level sentiment analysis; opinion mining; review analysis.

1. Introduction

Sentiment analysis can be defined as the analysis of text using natural language processing (NLP) and computational linguistics to extract subjective information from a given text source. It is the process of discovering as to whether a particular document, paragraph or sentence contains positive, negative or neutral opinion (Pang and Lee, 2008). There are three different levels of sentiment analysis — document level, sentence level and aspect, entity or feature level. In the document-level sentiment analysis, the aim is to identify the sentiment of the whole document, wherein it is assumed that each document contains opinion about a single entity. In the sentence-level sentiment analysis, the focus is to determine whether each sentence in a document contains positive, negative or neutral opinion, with the assumption that each sentence expresses opinion on only one topic. In the document-level or sentence-level sentiment analysis, the focus is more towards identifying positive, negative or neutral sentiments conveyed in the document or the sentence, and not at discovering as to "what" features of an entity are liked or disliked. The aspect-level sentiment analysis emphasises on identifying the features/aspects which are liked or disliked.

Sentiment analysis can be conducted using either machine learning classifiers or lexicons. Lexicons are a collection of words, tagged with opinions. Lexicon can be of two types: the domain-based lexicon and the general/basic lexicon. It has been observed that some words have different opinions across diverse domains (Hamilton *et al.*, 2016). As an illustration, the word "soft" expresses positive opinion in toy domain as in the case of "this toy is very soft", while it conveys negative opinion in sports domain as in "He is a soft hockey player". The domain-dependent words are tagged with different opinions across various domain lexicons depending on their connotation in the concerned domain, while general lexicon considers only one opinion for every word (Pimentel *et al.*, 2011). Sentiment analysis performance thus may vary depending on the lexicon used for sentiment classification.

In this paper, we study the effectiveness of domain lexicons in sentiment analysis by considering three different lexicons, one each from the domains of car, guitar and book. This paper contributes to the literature by comparing, using a quantitative metric, the effectiveness of domain-based lexicons vis-à-vis general lexicon when used for aspect-level sentiment analysis. It also investigates the impact of the use of domain-based lexicons in conjunction with general lexicon as compared to that of using individually the domain-based lexicons or general lexicon.

The remaining sections of this paper is organised as follows: Section 2 presents a brief review of the state of the art in sentiment analysis research; Sec. 3 describes the methodology adopted for the study and the data used for the purpose; the results and analysis are elaborated in Sec. 4; and the conclusion is presented in Sec. 5.

2. Literature Review

Sentiment analysis is a multidisciplinary task and encompasses NLP, information retrieval, information extraction, text mining, computational linguistics, psychology and predictive analysis (Chandrasekaran and Vinodhini, 2012). Sentiment analysis has various applications, like understanding consumers' purchasing behaviour, determining sellers' marketing strategy or manufacturers' production and design strategy, etc. (Chandrasekaran and Vinodhini, 2012; Vinodhini and Chandrasekaran, 2017; Nabareseh *et al.*, 2018). It can also be used in hotspot detection (Li and Wu, 2010), study of dynamic events like political sentiment (Ebrahimi *et al.*, 2017) and financial markets (Chan and Chong, 2017). Investors' decisions in financial markets are likely to be influenced by media sentiments. Busse and Green (2002) and Heston and Sinha (2016) found that a firm's stock price increases significantly within 15 s of CNBC airing a positive report about that firm. Sentiment analysis can be very useful in identifying the factors responsible for negative reviews (Valdivia *et al.*, 2017), and can be used for enhancing the capabilities of recommender systems and customer relationship management tools (Cambria, 2016).

The analysis can be performed at three different levels — document level, sentence level and aspect level. In the document-level analysis, a particular document is divided into chunks and sentiment contributions of these chunks are used to decide upon the final sentiment it conveys (Farra *et al.*, 2010). Another approach to document-level sentiment analysis is through clustering, wherein Term Frequency–Inverse Document Frequency (TF–IDF) weighing scheme can be used for the computation of k-means to cluster a document into positive and negative chunks (Li and Liu, 2010; Al-Sharuee *et al.*, 2018). TF–IDF technique expresses the quality of rare terms which have higher discrimination power as compared to the frequent terms. One major problem with document-level sentiment analysis is that a document may contain several contradictory opinions (Farra *et al.*, 2010). To handle this problem, hierarchal classification approach can be used wherein sentence-level sentiment classification is first undertaken; the classified sentences are then used as the input for document-level sentiment classification (Moraes *et al.*, 2013). Sentence-level sentiment analysis endeavours to identify opinion contained in each subjective sentence (Farra *et al.*, 2010; Appel *et al.*, 2016; Abdi *et al.*, 2018).

Both document- and sentence-level sentiment analyses focus on identifying the overall sentiment of the document or the sentence, but do not ascertain "what" features of the entity are liked or disliked by the opinion providers. This issue is handled in the aspect-level sentiment analysis (Che *et al.*, 2015; Pham and Le, 2018). Aspect-level sentiment analysis has three main steps — aspect identification or extraction, sentiment classification and aggregation or summarisation (Zhang *et al.*, 2010; Chandrasekaran and Vinodhini, 2012; Schouten and Frasincar, 2016; Hu *et al.*, 2017). Aspects can be extracted using frequent nouns or noun phrases (Zha *et al.*, 2014). Khan *et al.* (2010) made a case for considering auxiliary verb present in the sentence to improve sentiment classification as they found that about 82% of the aspects and 85% of opinion-oriented sentences have auxiliary verbs.

Word representations play a significant role in sentiment analysis performances. Word embedding can be defined as d-dimensional space representations of words, encoded as dense numerical vectors (Rojas-Barahona, 2016; Jianqiang *et al.*, 2018). Word embedding is used for word representation in low-dimensional distribution and can be generated using Extreme Learning Machine (ELM). Bag-of-Words is another widely used approach for word representation (Lauren *et al.*, 2018; Rudkowsky *et al.*, 2018). Word representation using document-level sentiment ratio on target (DLJT) was found to significantly improve sentiment analysis performance (Li *et al.*, 2017).

Sentiment analysis can be done using either machine learning or lexicon-based techniques (Zheng and Ye, 2009; Smatana *et al.*, 2013; Anami *et al.*, 2014; Jurek *et al.*, 2015), with the machine learning approach being more accurate but slow when compared to lexicon-based approach (Ravi and Ravi, 2015). Thus, for real-time applications, lexicon-based approach is preferred over machine learning approach (Chaovalit and Zhou, 2005; Vinodhini and Chandrasekaran, 2017; Dey *et al.*, 2018). In using machine learning techniques, one can use either supervised or unsupervised methods (Medhat *et al.*, 2014; Ravi and Ravi, 2015). In supervised machine learning approach, text classification algorithms like SVM classifier (Kianmehr *et al.*, 2007; Zheng and Ye, 2009; Shi and Li, 2011; Dehkharghani *et al.*, 2012; Raut and Londhe, 2014; Okada *et al.*, 2014; Abdi *et al.*, 2018), naive Bayes (NB) classifier (Ghorpade and Ragha, 2012; Smatana *et al.*, 2013; Raut and Londhe, 2014), decision tree classifier (Raut and Londhe, 2014), neural network classifier (Kianmehr *et al.*, 2007; Dehkharghani *et al.*, 2012; Vinodhini and Chandrasekaran, 2016), convolutional neural network (Lee *et al.*, 2018; Ghiassi and Lee, 2018) and logistic regression classifier (Dehkharghani *et al.*, 2012) have been used. Vilares *et al.* (2017) used supervised method for sentiment analysis for reviews that contain words from more than one languages. To train the classifiers for supervised learning, training dataset with tagged opinion is required; opinion tagging can be done manually, automatically or by using a hybrid of these methods (Shi and Li, 2011; Dehkharghani *et al.*, 2012). Being a time-consuming process, the requirement of option-tagged data to train classifiers is a major drawback of supervised approach (Ravi and Ravi, 2015). This limitation is addressed in unsupervised techniques, which are implemented using clustering (Hadano *et al.*, 2011; García-Pablos *et al.*, 2018) techniques like k-means clustering (Hu *et al.*, 2013; Al-Sharuee *et al.*, 2018), spectral clustering (Unnisa and Raziuddin, 2016) or clustering using hieratical self-organising map (Albertini *et al.*, 2014).

Deep learning techniques, which do not require high level of feature engineering, have also been effectively used for aspect extraction (Araque *et al.*, 2017; Cambria *et al.*, 2017; Dohaiha *et al.*, 2018; Ma *et al.*, 2018; Al-Smadi *et al.*, 2018). Poria *et al.* (2016) used deep convocational neural network for this purpose. Goldberg (2016) used long short-term memory (LSTM) to improve the back-propagation process thereby enhancing aspect extraction performance of deep learning networks. LSTM uses gates to control the read, write and reset operations (Chen *et al.*, 2017). Ma *et al.* (2018) introduced Sentic LSTM wherein they added an extra output to insert token-level memory and concept-level inputs to improve the aspect extraction efficiency.

As mentioned earlier, sentiment analysis can also be undertaken using lexicons. Lexicon is a collection of opinion words, opinion phrases and idioms (Ding *et al.*, 2008; Hamouda, 2011). It is a dictionary of terms with corresponding sentiment scores (Muhammad *et al.*, 2013). A simple lexical entry is called a polar atom and is the minimum human understandable syntactic structure specifying polarity. A polar atom can be obtained using context coherency (Kanayama and Nasukawa, 2006). A polarity-tagged corpus contains phrases, sentences or documents tagged with their semantic orientation of being positive, negative or neutral. Lexicons are of two types, general lexicons and domain lexicons (Ding *et al.*, 2008). SentiWordNet is an example of general lexicon and it is openly available for research purposes. It assigns positive, negative and objective polarity scores to each synset of WordNet (Esuli *et al.*, 2006; Baccianella *et al.*, 2010; Philander and Zhong, 2016). English words bear different meanings based on the communities or domains they are used in (Hamilton *et al.*, 2016; Khan *et al.*, 2014) and these are taken care in domain lexicons dedicated to a given domain. The domain lexicons across different domains may thus have different sentiment scores for a given word (Dehkharghani *et al.*, 2012), unlike the general lexicons.

A lexicon can be developed using small seed words by corpus-based or dictionary-based approach. The corpus-based approach uses a collection of documents while the dictionary-based approach uses a machine-readable dictionary like WordNet (Dorr *et al.*, 2002; Muhammad *et al.*, 2013). A lexicon can be created using manual tagging, which is not scalable, or can be created using review sites which already have tagged reviews in terms of star ratings or sentences divided into "pros" and "cons" sections. However, domain-specific tags are not always present in such reviews (Kaji and Kitsuregawa, 2006; Muhammad *et al.*, 2013). Polarity-tagged corpus can also be built from HTML document by utilising the indicators such as certain layout structures or patterns (Kaji and Kitsuregawa, 2006). The widely used SentiWordNet lexicon has been created using dictionary-based approach (Guessoum and Zantout, 2001; Dehkharghani *et al.*, 2012; Ansari, 2015; Bucur, 2015; Che *et al.*, 2015). Distant-supervision is another way to create domain-focused lexicons (Muhammad *et al.*, 2013). Li *et al.* (2018) proposed a way for new word detection called domain-specific new words detection and word propagation (DWWP) system consisting of two parts: DW (words detection) that detects user-invented words, multiword expressions or converted words using manually created seed word list; and WP (word propagation) which is repeatedly performed to achieve convergence condition. Deng *et al.* (2017) combined multiple general lexicons to generate domain-specific lexicons. Sentiment analysis can also be performed using OntoSenticNet which is based on ontology and depends on the implicit meaning of the words associated with the concepts and does not blindly use keywords. It can combine multiple words to relate them with the concepts (Dragoni *et al.*, 2018).

Zhang *et al.* (2015) used both lexicon and machine learning approaches for sentiment analysis. They used general lexicon-based approach to create the training data and used it to train the machine learning classifiers. Musto *et al.* (2014) compared different available general lexicons like SentiWordNet, WordNet-Affect, MPQA and SenticNet, and found that SentiWordNet and MPQA performed better. Kim *et al.* (2016) had compared various domain-based lexicon development methods. Mao *et al.* (2015) compared domain lexicon and general lexicon using Chinese language reviews. They compared the performance when domain and general lexicons are used separately in document-level sentiment analysis. Muhammad *et al.* (2013) have used domain and general lexicons together to study the performance of sentiment analysis.

From the literature review, we can infer that no study has been undertaken to understand the performance of lexicon-based aspect-level sentiment analysis in English language reviews involving both domain and general lexicons together, with the domain lexicon being given the prominence in the case of same words appearing in both the lexicons. General lexicons tend to be rich as they are intended to cover a wide gamut of words. Moreover, as certain words have different meanings in different domains, some opinion wearing words also have different opinions in different domains; this calls for the use of domain lexicon. Given these, use of domain lexicons and general lexicons together, instead of using them separately, may increase the aspect-level sentiment analysis performance. This study is undertaken to address this gap in the literature, and has used the domain lexicons in the domains of car, guitar and book along with the SentiWordNet general lexicon for aspect-level sentiment analysis of related English language topical reviews. This study is, however, inspired by the works done by Mao *et al.* (2015) and Muhammad *et al.* (2013). Yet, the current study is different from the former, as it looks at the English language reviews, and is distinct from that of Muhammad *et al.* (2013), as they have given the same weightage to these lexicons if a term appeared in both the lexicons, while this study gives preference to domain lexicon in such cases.

This study envisages to contribute to the literature in terms of a quantitative comparison of performance in aspect-level sentiment analysis using both domain and general lexicons in English language, separately as well as together, and trusting the domain lexicon over the general lexicon in case a term appears in both the lexicons.

3. Methodology and Data Source

In this paper, we considered three domain-based lexicons along with the general lexicon SentiWordNet. These domain-based lexicons[1] are related to the domains of cars, books and guitars. Each domain lexicon considered has 4980 words along with their opinion scores. Table 1 presents the details about these three domain lexicons.

Table 1. Details of positive and negative sentiment words in each lexicon.

Lexicon domain	Positive sentiment words	Negative sentiment words
Car	2145	2835
Book	2642	2338
Guitar	2412	2568

Source: Made by the authors.

[1] https://nlp.stanford.edu/projects/socialsent/.

Table 2. Reviews considered for the corpora.

Domain	Total review	Brand/Title/Model	Source
Car	100	• Maruti Suzuki Dzire • Hyundai i20 • Tata Scorpio • Honda City • Maruti Alto	Gaadi.com
Book	100	• Think and Grow Rich • Adolf Hitler • The Power of Your Subconscious Mind • Sita: Warrior of Mithila • GK-2018	Amazon.in
Guitar	100	• 38" Black Acoustic Guitar • Fender Acoustic Guitar • 41 Inch Full Size Black Handcrafted • ZENY 38" New Beginners Acoustic Guitar • Crescent MG38-BK 38" Acoustic Guitar	Amazon.in

Source: Made by the authors.

The study of performance of these lexicons requires related corpora. These corpora are taken from the domain-based reviews, extracted from relevant sites. For reviews related to cars, 100 reviews pertaining to five different car brands, 20 for each brand, are randomly picked up from the site gaadi.com. This is an automobile website in India, and offers its users with research on various car models, buying guide, selling platform, etc., in addition to hosting user reviews related to cars. The reviews for books and guitars are taken from the e-commerce site amazon.in, which is the Indian portal of the e-commerce giant Amazon.com. For the book domain, 20 reviews each for five different books totalling 100 reviews are randomly collected, and similarly for guitar, 20 reviews of five different guitar models totalling 100 are randomly taken. The aforesaid reviews were gathered from the corresponding sites using the Octoparse[2] tool, which is an open-source tool and is freely available for data collection. Details about selected reviews of all three domains are listed in Table 2.

Aspect-level sentiment analysis is performed for all three domains — car, book and guitar. At first, aspects are extracted separately for each case using the hybrid approach of aspect extraction. In this approach, first, the frequent nouns/noun phrases are extracted using the part-of-speech (POS) tagging. To select infrequent nouns/noun phrases, we used adjectival relationship between noun and opinion words. First opinion words related to frequent nouns are extracted, and then these opinion words are used to extract infrequent nouns. Final list of aspects includes both frequent nouns/noun phrases and infrequent nouns.

For each of these cases, aspect-level sentiment analysis is performed using the corresponding domain lexicon, the SentiWordNet lexicon and then using both these

[2] https://www.octoparse.com/.

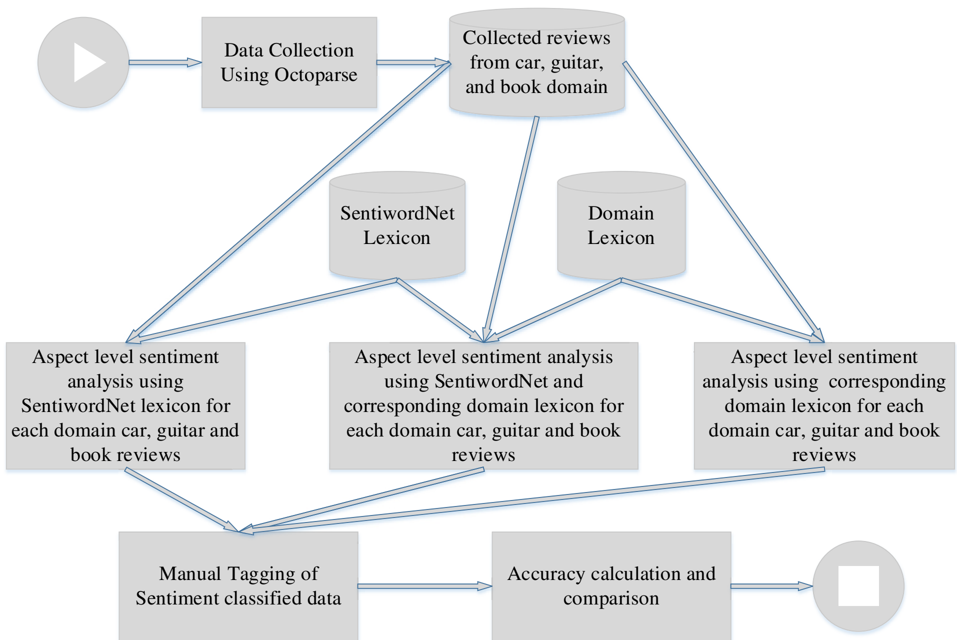
lexicons simultaneously. Aspect-wise sentiment scores for each review are calculated using the formula presented in Eq. (1):

$$\mathrm{score}(a,r)=\sum_{s\epsilon r}\left(\sum_{t\epsilon s}\left(\frac{\mathrm{sco}(t,d)}{\mathrm{distance}(a,t)}\right)\right), \tag{1}$$

where score(a,r) is the score of an aspect a in review r; s represents a sentence in the review; t represents a term in a sentence; sco(t,d) is the sentiment score of a term t in the lexicon d; distance(a,t) represents the distance between aspect term a and term t in the considered sentence.

The sentiment classification accuracy (CA) achieved in each of these cases with the corresponding lexicon(s) is then compared. Flowchart of the process followed in this analysis is presented in Fig. 1.

The first step is *data collection using Octoparse* wherein we collect online reviews for car, guitar and book from gaadi.com and amazon.in websites using Octoparse tool. We then save the *collected reviews from car*, *guitar and book domains* as three datasets for the analysis. The *aspect-level sentiment analysis* involves three parallel processes wherein we perform the analysis individually on all the three datasets using (a) SentiWordNet alone, (b) domain lexicon alone and (c) SentiWordNet and domain lexicon together. In the last scenario of simultaneous use of domain lexicon and SentiWordNet, if the same word appears in both the lexicons, we use the opinion as



Source: Made by the authors.

Fig. 1. Flowchart of the proposed work.

Table 3. An example for aspect-wise manual tagging.

S. No.	Aspect	Sentiment score using SentiWordNet	Agree with manual tag	Sentiment score using car lexicon	Agree with manual tag
1	Dzire	0.01	T	0.03	T
2	Fuel economy	1.02	T	0.22	T
3	Drive	0.64	T	0.66	T
4	Look	−0.05	F	0.11	T
5	Seat	−0.35	T	−0.40	T
6	Music system	0.02	T	0.17	T
7	Value for money	0.21	T	0.21	T

Source: Made by the authors.

given in the domain lexicon over that in SentiWordNet. We consider the opinions of words from SentiWordNet in the cases where the concerned words are not present in the corresponding domain lexicon. The reason behind this is the understanding that domain lexicon is better suited for the sentiment analysis of corpora belonging to a particular domain, as compared to the general lexicon (Park *et al.*, 2015; Hamilton *et al.*, 2016). In the step *manual tagging of the sentiment classified data*, we manually tag the opinions expressed for the aspects; this is later used to compare with the opinions extracted through sentiment analysis. In the final step of *accuracy calculation and comparison*, we compute the sentiment CA as

$$\mathrm{CA} = \frac{T}{(T+F)}, \tag{2}$$

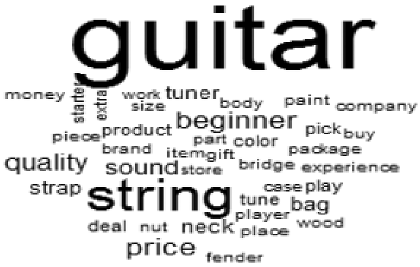
where T represents the number of times the sentiment is correctly predicted and F represents otherwise. The T and F values are computed based on the manual tagging of opinion undertaken in the penultimate step. Table 3 illustrates an example of manual tagging for a review in the car domain:

> "Dzire is easy to drive and is currently returning me an excellent fuel economy of 19.6 km/litre. The rear seat is uncomfortable for three people. The car looks average, the music system is alright. The car drives well. It's a true value for money product."

4. Results and Analysis

The top 50 aspects for each of the three domains, as extracted from the aspect-level sentiment analysis, are presented in the form of a wordcloud[3] in Figs. 2–4, respectively. In a wordcloud, the size of a given word is proportional to its frequency of appearance in the selected document. In Fig. 2 the word "guitar" has the largest size, implying that it appeared the maximum number of times in the selected reviews. The important aspects identified in this domain are guitar, string, quality, sound, price, beginner, bag, neck, tuner, etc.

[3] https://cran.r-project.org/web/packages/wordcloud/wordcloud.pdf.



Source: Made by the authors.

Fig. 2. Wordcloud for the aspects in guitar domain.

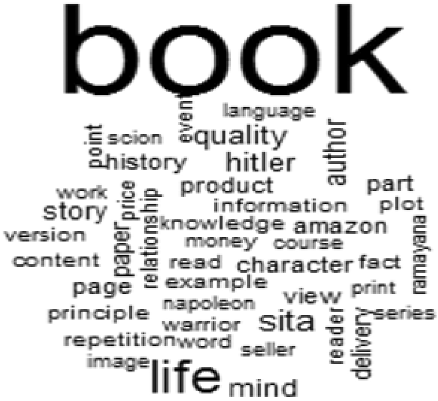
Source: Made by the authors.

Fig. 3. Wordcloud for the aspects in car domain.

In the case of car domain, major identified aspects are car, service, vehicle, engine, road, diesel, honda, scorpio, alto, fuel, etc., as presented in Fig. 3 in the form of wordcloud.

Similarly, in the case of book domain, the major identified aspects are book, life, author, quality, sita, hitler, story, history, mind, page, principle, plot, part, etc., as presented in the wordcloud in Fig. 4.

After aspect extraction, the sentiment classification is performed. In the case of car-domain reviews, aspect-level sentiment analysis carried out using SentiWordNet resulted in 439 instances of correct classification in terms of positive or negative sentiments, while in 187 instances the sentiments were incorrectly classified. The classification performance improved slightly when the car-domain lexicon was used instead of SentiWordNet — 458 instances of correct classification, while 168 instances of wrong classification. However, the performance was best when both the car-domain lexicon and SentiWordNet were used simultaneously — 495 instances of correct classification and a miss in the case of 131 instances. The details are presented in Table 4.



Source: Made by the authors.

Fig. 4. Wordcloud for the aspects in book domain.

Table 4. Car-domain sentiment analysis.

Lexicon(s) used	True classification	False classification	Accuracy (%)
SentiWordNet	439	187	70.13
Car lexicon	458	168	73.16
Car+SentiWordNet lexicons	495	131	79.07

Source: Made by the authors.

In the case of book-domain reviews, SentiWordNet lexicon resulted in 439 instances of correct sentiment classification of the aspects and 187 instances of incorrect classification. With only book-domain lexicon, there was however only 196 instances of correct classification with 79 instances of incorrect classification. The best classification performance in this case too was found with the simultaneous use of book lexicon along with SentiWordNet, resulting in 224 instances of correct classification with a miss in 51 instances. The details are presented in Table 5.

A similar classification performance in aspect-level sentiment analysis was evaluated for the guitar domain and the details are presented in Table 6.

When the aspect-level sentiment classification accuracies obtained with simultaneous use of domain lexicon and SentiWordNet lexicon are compared against those obtained using SentiWordNet lexicon or the corresponding domain lexicons individually, the classification performance is seen to be the best in the case of former in all the three scenarios, as plotted in Fig. 5. However, as seen in the case of book domain, the performance of the domain-specific lexicon, when used alone, is lesser than that of general lexicon. The analysis carried out brings to the fore that the use of domain lexicon may not always result in better performance over a general

Table 5. Book-domain sentiment analysis.

Lexicon(s) used	True classification	False classification	Accuracy (%)
SentiWordNet	213	62	77.45
Book lexicon	196	79	71.27
Book+SentiWordNet lexicons	224	51	81.45

Source: Made by the authors.

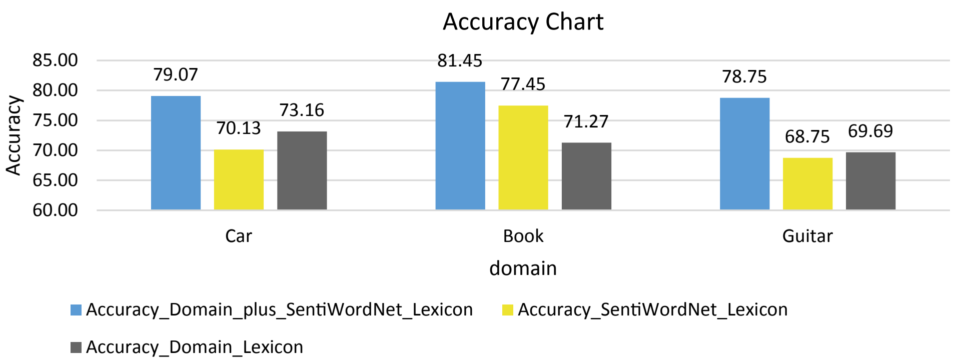
Table 6. Guitar-domain sentiment analysis.

Lexicon(s) used	True classification	False classification	Accuracy (%)
SentiWordNet	220	100	68.75
Guitar lexicon	223	97	69.69
Guitar+SentiWordNet lexicons	252	68	78.75

Source: Made by the authors.

lexicon, as seems to be the current understanding (Park *et al.*, 2015; Hamilton *et al.*, 2016).

The reason for the less than expected performance of domain lexicon in the case of books is the general nature of book reviews as compared to those of cars and guitars. The book reviews considered here were taken from those of customers as appearing in Amazon.in, and not of critics. Thus, the words used were very generic, and SentiWordNet is richer as compared to the book lexicon for such words. Some of the opinion wearing words used in the reviews are not present in domain lexicon, which led to zero score for those words while using the domain lexicon, thereby lowering the classification accuracy. Thus, as brought out from this analysis, the performance of domain lexicons depends on the content of reviews in terms of usage



Source: Made by the authors.

Fig. 5. Aspect-level sentiment analysis classification performance based on lexicons used.

of domain-specific words rather than common English words having the same opinions across domains, as well as the richness of the domain lexicon. While it is intuitive that domain lexicon will perform better than general lexicon, we found that general lexicon can perform better than domain lexicon if the concerned corpora contain more non-technical terms as compared to technical terms, and thus the performance depends on the content of the corpora. However, simultaneous use of domain-specific and general lexicons irons out such limitations and results in better performance across the domains. Dey *et al.* (2018) used general lexicon for sentiment analysis on different datasets but did not perform aspect-level sentiment analysis as we have done; the performance in book domain they reported was 68%. Similarly, the proposed approach performs better than the approaches used in Li *et al.* (2017) and García-Pablos *et al.* (2018). However, it is important to note that the datasets and sentiment analysis techniques used in the previous studies and current study are different, and thus are not amenable for direct comparison.

5. Conclusion

In this paper, we study the effectiveness of domain lexicons versus general lexicons in aspect-level sentiment analysis. In the existing literature, we did not come across any such quantitative comparison of classification accuracy for aspect-level sentiment analysis undertaken using various lexicons — domain-based and general, used independently and simultaneously. This paper is thus an endeavour to contribute with such a performance analysis to the literature.

The current analysis concludes that the best classification performance is achieved when the domain-specific lexicon is used along with a general lexicon (SentiWordNet, in this case), and not when any of these are used independently. The study thus underscores the importance of general lexicon even in the presence of domain-specific lexicons. It also highlights that effectiveness of domain lexicons depends on the nature of words contained in the corpora.

In future, we propose to conduct analysis of aspect-level sentiment analysis with the simultaneous use of domain and general lexicons, wherein we assign different weights to the scores of a word appearing in both lexicons, and evaluate if such weights improve on the classification performance. This study also calls for development of domain lexicons across domains, as the use of these along with general lexicon is found to improve the sentiment classification performance.

References

Abdi, A, SM Shamsuddin, S Hasan and J Piran (2018). Machine learning-based multi-documents sentiment-oriented summarization using linguistic treatment. *Expert Systems with Applications*, 109, 66–85.

Albertini, S, A Zamberletti and I Gallo (2014). Unsupervised feature learning for sentiment classification of short documents. *Journal for Language Technology and Computational Linguistics*, 29(1), 1–15.

Al-Smadi, M, O Qawasmeh, M Al-Ayyoub, Y Jararweh and B Gupta (2018). Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews. *Journal of Computational Science*, 27, 386–393.

Al-Sharuee, MT, F Liu and M Pratama (2018). Sentiment analysis: An automatic contextual analysis and ensemble clustering approach and comparison. *Data & Knowledge Engineering*, 115, 194–213.

Anami, BS, RS Wadawadagi and VB Pagi (2014). Machine learning techniques in Web content mining: A comparative analysis. *Journal of Information & Knowledge Management*, 13(1), 1450005-1–1450005-12.

Ansari, D (2015). Sentiment polarity classification using structural features. In *Proceedings of the 2015 IEEE International Conference on Data Mining Workshop*, Atlantic City, NJ, November, pp. 1270–1273.

Appel, O, F Chiclana, J Carter and H Fujita (2016). A hybrid approach to the sentiment analysis problem at the sentence level. *Knowledge-Based Systems*, 108, 110–124.

Araque, O, I Corcuera-Platas, JF Sánchez-Rada and CA Iglesias (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications. *Expert Systems with Applications*, 77, 236–246.

Baccianella, S, A Esuli and F Sebastiani (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation*, Valletta, Malta, May, pp. 2200–2204.

Bucur, C (2015). Using opinion mining techniques in tourism. *Procedia Economics and Finance*, 23, 1666–1673.

Cambria, E (2016). Affective computing and sentiment analysis. *IEEE Intelligent Systems*, 31(1), 102–107.

Cambria, E, S Poria, A Gelbukh and M Thelwall (2017). Sentiment analysis is a big suitcase. *IEEE Intelligent Systems*, 32(6), 74–80.

Chandrasekaran, RM and G Vinodhini (2012). Sentiment analysis and opinion mining: A survey. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(6), 282–292.

Chan, SW and MW Chong (2017). Sentiment analysis in financial texts. *Decision Support Systems*, 94, 53–64.

Chaovalit, P and L Zhou (2005). Movie review mining: A comparison between supervised and unsupervised classification approaches. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, Big Island, HI, January, pp. 1–9.

Che, W, Y Zhao, H Guo, Z Su and T Liu (2015). Sentence compression for aspect based sentiment analysis. *IEEE/ACM Transactions on Speech and Language Processing*, 23(12), 2111–2124.

Chen, T, R Xu, Y He and X Wang (2017). Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. *Expert Systems with Applications*, 72, 221–230.

Chen, YS, LH Chen and Y Takama (2015). Proposal of LDA-based sentiment visualization of hotel reviews. In *Proceedings of the 2015 IEEE International Conference on Data Mining Workshop*, Atlantic City, NJ, November, pp. 687–693.

Dehkharghani, R, B Yanikoglu, D Tapucu and Y Saygin (2012). Adaptation and use of subjectivity lexicons for domain dependent sentiment classification. In *Proceedings of the IEEE 12th International Conference on Data Mining Workshops Adaptation*, Brussels, December, pp. 669–673.

Deng, S, AP Sinha and H Zhao (2017). Adapting sentiment lexicons to domain-specific social media texts. *Decision Support Systems*, 94, 65–76.

Dey, A, M Jenamani and JJ Thakkar (2018). *Senti-N-Gram*: An n-gram lexicon for sentiment analysis. *Expert Systems with Applications*, 103, 92–105.

Ding, X, B Liu and PS Yu (2008). A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, Palo Alto, February, pp. 231–240.

Dohaiha, HH, PWC Prasad, A Maag and A Alsadoon (2018). Deep learning for aspect-based sentiment analysis: A comparative review. *Expert Systems with Applications*, 118, 272–299.

Dorr, BJ, GA Levow and D Lin (2002). Construction of a Chinese–English verb lexicon for machine translation and embedded multilingual applications. *Machine Translation*, 17(1), 99–137.

Dragoni, M, S Poria and E Cambria (2018). OntoSenticNet: A commonsense ontology for sentiment analysis. *IEEE Intelligent Systems*, 33(3), 77–85.

Ebrahimi, M, AH Yazdavar and A Sheth (2017). Challenges of sentiment analysis for dynamic events. *IEEE Intelligent Systems*, 32(5), 70–75.

Esuli, A, F Sebastiani and VG Moruzzi (2006). SENTIWORD NET?: A publicly available lexical resource for opinion mining. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, Genoa, May, pp. 417–422.

Farra, N, E Challita, RA Assi and H Hajj (2010). Sentence-level and document-level sentiment mining for Arabic texts. In *Proceedings of the 2010 IEEE International Conference on Data Mining Workshops*, Sydney, December, pp. 1114–1119.

García-Pablos, A, M Cuadros and G Rigau (2018). W2VLDA: Almost unsupervised system for aspect based sentiment analysis. *Expert Systems with Applications*, 91, 127–137.

Ghiassi, M and S Lee (2018). A domain transferable lexicon set for Twitter sentiment analysis using a supervised machine learning approach. *Expert Systems with Applications*, 106, 197–216.

Ghorpade, T and L Ragha (2012). Featured based sentiment classification for hotel reviews using NLP and Bayesian classification. In *Proceedings of the 2012 International Conference on Communication, Information & Computing Technology*, Mumbai, October, pp. 1–5.

Goldberg, Y (2016). A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research*, 57, 345–420.

Guessoum, A and R Zantout (2001). A methodology for a semi-automatic evaluation of the lexicons of machine translation systems. *Machine Translation*, 16(1), 127–149.

Hadano, M, K Shimada and T Endo (2011). Aspect identification of sentiment sentences using a clustering algorithm. *Procedia — Social and Behavioral Sciences*, 27, 22–31.

Hamilton, WL, K Clark, J Leskovec and D Jurafsky (2016). Inducing domain-specific sentiment lexicons from unlabeled corpora. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Austin, November, pp. 595–605.

Hamouda, A (2011). Reviews classification using SentiWordNet lexicon. *The Online Journal on Computer Science and Information Technology*, 2(1), 120–123.

Heston, SL and NR Sinha (2016). News vs. sentiment: Predicting stock returns from news stories. *Financial Analysts Journal*, 73(3), 67–83.

Hu, X, J Tang, H Gao and H Liu (2013). Unsupervised sentiment analysis with emotional signals. In *Proceedings of the 22nd International Conference on World Wide Web*, Rio de Janeiro, May, pp. 607–618.

Hu, YH, YL Chen and HL Chou (2017). Opinion mining from online hotel reviews: A text summarization approach. *Information Processing & Management*, 53(1), 436–449.

Jianqiang, Z, G Xiaolin and Z Xuejun (2018). Deep convolution neural networks for Twitter sentiment analysis. *IEEE Access*, 6, 23253–23260.

Jurek, A, MD Mulvenna and Y Bi (2015). Improved lexicon-based sentiment analysis for social media analytics. *Security Informatics*, 4(1), 9-1–9-13.

Kaji, N and M Kitsuregawa (2006). Automatic construction of polarity-tagged corpus from HTML documents. In *Proceedings of the COLING/ACL on Main Conference Poster Sessions*, Sydney, July, pp. 452–459.

Kanayama, H and T Nasukawa (2006). Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, Sydney, July, pp. 355–363.

Kianmehr, K, H Zhang, K Nikolov, T Özyer and R Alhajj (2007). Utilising neural network and support vector machine for gene expression classification. *Journal of Information & Knowledge Management*, 6(4), 251–260.

Kim, MS, JW Kim and C Jing (2016). Comparison of domain-specific lexicon construction methods for sentiment analysis. *Advanced Science and Technology Letters*, 135, 152–156.

Khan, K, B Baharudin and A Khan (2010). Automatic extraction of features and opinion oriented sentences from customer reviews. *World Academy of Science, Engineering and Technology: International Journal of Social, Behavioural, Educational, Economic, Business and Industrial Engineering*, 4(1), 102–106.

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

Lauren, P, G Qu, J Yang, P Watta, GB Huang and A Lendasse (2018). Generating word embeddings from an extreme learning machine for sentiment analysis and sequence labeling tasks. *Cognitive Computation*, 10(4), 625–638.

Lee, G, J Jeong, S Seo, C Kim and P Kang (2018). Sentiment classification with word localization based on weakly supervised learning with a convolutional neural network. *Knowledge-Based Systems*, 152, 70–82.

Li, G and F Liu (2010). A clustering-based approach on sentiment analysis. In *Proceedings of the 2010 International Conference on Intelligent Systems and Knowledge Engineering*, Hangzhou, November, pp. 331–337.

Li, N and D Wu (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems*, 48(1), 354–368.

Li, W, K Guo, Y Shi, L Zhu and Y Zheng (2018). DWWP: Domain-specific new words detection and word propagation system for sentiment analysis in the tourism domain. *Knowledge-Based Systems*, 146, 203–214.

Li, Y, Q Pan, T Yang, S Wang, J Tang and E Cambria (2017). Learning word representations for sentiment analysis. *Cognitive Computation*, 9(6), 843–851.

Ma, Y, H Peng, T Khan, E Cambria and A Hussain (2018). Sentic LSTM: A hybrid network for targeted aspect-based sentiment analysis. *Cognitive Computation*, 10(4), 639–650.

Mao, K, J Niu, X Wang, L Wang and M Qiu (2015). Cross-domain sentiment analysis of product reviews by combining lexicon-based and learn-based techniques. In *Proceedings of the IEEE 17th International Conference on High Performance Computing and Communications*, New York, August, pp. 351–356.

Medhat, W, A Hassan and H Korashy (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113.

Moraes, R, JF Valiati and WP Gavião Neto (2013). Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications*, 40(1), 621–633.

Muhammad, A, N Wiratunga, R Lothian and R Glassey (2013). Domain-based lexicon enhancement for sentiment analysis. In *Proceedings of the BCS-SGAI Workshop on Social Media Analysis*, pp. 7–18.

Musto, C, G Semeraro and M Polignano (2014). A comparison of lexicon-based approaches for sentiment analysis of microblog. *Information Filtering and Retrieval*, 1314, 59–68.

Nabareseh, S, E Afful-Dadzie and P Klimek (2018). Leveraging fine-grained sentiment analysis for competitivity. *Journal of Information & Knowledge Management*, 17(1), 1850018-1–1850018-20.

Okada, M, K Takeuchi and K Hashimoto (2014). A method to classify customer reviews of Japanese hotels by support vector machine using estimation sentence patterns information. In *Proceedings of the 2014 IIAI 3rd International Conference on Advanced Applied Informatics*, pp. 567–572.

Pang, B and L Lee (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135.

Park, S, W Lee and IC Moon (2015). Efficient extraction of domain specific sentiment lexicon with active learning. *Pattern Recognition Letters*, 56, 38–44.

Pham, DH and AC Le (2018). Learning multiple layers of knowledge representation for aspect based sentiment analysis. *Data & Knowledge Engineering*, 114, 26–39.

Philander, K and Y Zhong (2016). Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *International Journal of Hospitality Management*, 55, 16–24.

Pimentel, J, MC L'Homme and MÈ Laneville (2011). General and specialized lexical resources: A study on the potential of combining efforts to enrich formal lexicons. *International Journal of Lexicography*, 25(1), 152–190.

Poria, S, E Cambria and A Gelbukh (2016). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, 108, 42–49.

Raut, VB and DD Londhe (2014). Opinion mining and summarization of hotel reviews. In *Proceedings of the 2014 6th International Conference on Computational Intelligence and Communication Networks*, Bhopal, November, pp. 556–559.

Ravi, K and V Ravi (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*, 89, 14–46.

Rojas-Barahona, LM (2016). Deep learning for sentiment analysis. *Language and Linguistics Compass*, 10(12), 701–719.

Rudkowsky, E, M Haselmayer, M Wastian, M Jenny, Š Emrich and M Sedlmair (2018). More than bags of words: Sentiment analysis with word embeddings. *Communication Methods and Measures*, 12(2–3), 140–157.

Schouten, K and F Frasincar (2016). Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering*, 28(3), 813–830.

Shi, H and X Li (2011). A sentiment analysis model for hotel reviews based on supervised learning. In *Proceedings of the 2011 International Conference on Machine Learning and Cybernetics*, Guilin, pp. 950–954.

Smatana, M, P Koncz, P Smatana and J Paralic (2013). Active learning enhanced semi-automatic annotation tool for aspect-based sentiment analysis. In *Proceedings of the 2013 IEEE 11th International Symposium on Intelligent Systems and Informatics*, Subotica, September, pp. 191–194.

Unnisa, M and S Raziuddin (2016). Opinion mining on Twitter data using unsupervised learning technique. *International Journal of Computer Applications*, 148(12), 12–19.

Valdivia, A, MV Luzón and F Herrera (2017). Sentiment analysis in TripAdvisor. *IEEE Intelligent Systems*, 32(4), 72–77.

Vilares, D, MA Alonso and C Gómez-Rodríguez (2017). Supervised sentiment analysis in multilingual environments. *Information Processing & Management*, 53(3), 595–607.

Vinodhini, G and RM Chandrasekaran (2016). A comparative performance evaluation of neural network based approach for sentiment classification of online reviews. *Journal of King Saud University — Computer and Information Sciences*, 28(1), 2–12.

Vinodhini, G and RM Chandrasekaran (2017). A sampling based sentiment mining approach for e-commerce applications. *Information Processing & Management*, 53(1), 223–236.

Zha, ZJ, J Yu, J Tang, M Wang and TS Chua (2014). Product aspect ranking and its applications. *IEEE Transactions on Knowledge and Data Engineering*, 26(5), 1211–1224.

Zhang, K, R Narayanan and A Choudhary (2010). Voice of the customers: Mining online customer reviews for product feature-based ranking. In *Proceedings of the 3rd Wonference on Online Social Networks*, Boston, June, pp. 11–19.

Zhang, L, R Ghosh, M Dekhil, M Hsu and B Liu (2015). Combining lexicon-based and learning-based methods for Twitter sentiment analysis. Technical Report No. HPL-2011-89, HP Laboratories, pp. 1–7.

Zheng, W and Q Ye (2009). Sentiment classification of Chinese traveler reviews by support vector machine algorithm. In *Proceedings of the Third International Symposium on Intelligent Information Technology Application Sentiment*, Shanghai, November, pp. 335–338.