Cluster Comput
DOI 10.1007/s10586-017-1096-9

Aspect-based opinion mining framework using heuristic patterns

Muhammad Zubair Asghar[1] · Aurangzeb Khan[2] · Syeda Rabail Zahra[1] · Shakeel Ahmad[3] · Fazal Masud Kundi[1]

Received: 14 April 2017 / Revised: 21 July 2017 / Accepted: 1 August 2017

Abstract The aspect-based online opinions expressed by users on social media sites have become a popular source of information for consumers regarding their purchase decisions as well as for companies seeking opinions on their products. Therefore, it is important to develop aspect-based opinion mining applications with an emphasis on extracting and classifying the aspect-based opinions expressed by users about products in a given review. Previous studies have used a limited set of heuristic patterns for aspect extraction with both supervised (annotated-dataset-based) and unsupervised (lexical-resource-based) aspect-related sentiment classification algorithms. However, the present study proposes an integrated framework comprising of an extended set of heuristic patterns for aspect extraction, a hybrid sentiment classification module with the additional support of intensifiers and negations, and a summary generator. The performance evaluation of the proposed aspect-based opinion mining system using state-of-the-art methods shows that the proposed system outperforms the alternative methods in terms of better precision, recall and F-measure, since it achieves an average precision of 85%, an average recall of 73% and an average F-measure of 0.78. The comparative results indicate that the proposed technique provides more efficient results for the aspect-sentiment extraction, classification and summary generation of online product reviews.

✉ Muhammad Zubair Asghar
zubair@gu.edu.pk

1 Institute of Computing and Information Technology, Gomal University, D.I. Khan, Pakistan

2 Department of Computer Science, University of Science and Technology, Bannu, Pakistan

3 Faculty of Computing and Information Technology in Rabigh (FCITR), King Abdul Aziz University (KAU), Jeddah, Saudi Arabia

Keywords Aspect-based · Opinion mining · Sentiment analysis · Patterns · Hybrid · Corpus-based

1 Introduction

The rapid evolution of social media sites has prompted online users to share their opinions about policies, services, healthcare, politics and products [6]. Prior to this boom in online opinion sharing, when an individual needed to make a decision, he/she typically asked for opinions from friends, neighbours and family members. Similarly, when an organisation needed to obtain customers' feedback about its products and services, it conducted opinion polls, surveys and focus groups. With the emergence of Web 2.0, a rich source of opinions and facts expressed by people worldwide in relation to particular entities, including products, issues, policies and services, has become readily available. The increased use of social media sites had encouraged consumers to rely on the reviews available on such sites when making purchase decisions [15].

Opinion mining (OM) or sentiment analysis (SA) aims to develop applications that can assist individuals in making decisions regarding their purchases as well as help business organisations to determine the strengths and weaknesses of their products. In recent years, OM applications have increasingly gained the attention of both online users and researchers. The main focus of the research in this field has been on tasks, including emotion detection, opinion spam detection, concept and behaviour analysis, and aspect-based opinion mining [9,15,21]. However, due to the increasing interest of customers in knowing about the various aspects of products, aspect-based OM has emerged as an active area of research [21]. The basic aim of aspect-based OM is to detect aspects and identify the sentiments related to each aspect that are expressed by users in online reviews. In aspect-based OM, we therefore try to extract both aspects and the opinions expressed by users on the identified aspects of products. The basic tasks of aspect-based OM include: (i) extracting aspect terms, (ii) identification of sentiments expressed towards the extracted aspects, and (iii) visualisation of the extracted aspect-based opinion-related information [9].

In aspect-based OM, the term "aspect" is defined as the important features of products as rated by customers. Product reviews offer a mixture of positive and negative opinions about a given product's different aspects. For example, the sentence "I love my laptop's touchscreen" contains the positive opinion "love" regarding the aspect "touchscreen" of the entity "laptop". Similarly, the sentence "the battery life is not good" conveys the negative opinion "not good" about the aspect "battery life".

The principal focus of the present study is on extracting aspects, classifying aspect-related sentiments and generating an aspect-level summary. The proposed framework is motivated by previous studies [8,12,16] conducted on aspect-level OM. Those previous studies used a limited set of heuristic patterns for aspect extraction with supervised (annotated-dataset-based) and unsupervised (lexical-resource-based) aspect-related sentiment classification algorithms. However, this study proposes an integrated framework comprising of an extended set of linguistic patterns for aspect extraction, a hybrid sentiment classification module with the additional support of intensifiers and negations, and a user-friendly summary generator.

This study makes following key contributions:

- We propose an extended set of heuristic patterns for extracting aspects and their related sentiments as well as developing an algorithm for aspect-sentiment extraction.
- We propose a hybrid sentiment classification scheme for classifying aspect-related sentiments using lexicon-based and corpus-based classification techniques with the additional support of intensifiers and negations.
- We propose an aspect-sentiment-based summary with a user-friendly interface based on the extracted and classified aspect-related sentiments concerning a given product.

2 Related work

Aspect-level OM is a challenging task due to the unstructured nature of the reviews produced by the online community. Yet, several approaches have already been proposed for aspect-based opinion mining [8,18,20,22,23,27,31]. In aspect-level OM, three major tasks have been identified by researchers, namely (i) aspect extraction, (ii) aspect-related sentiment classification, and (iii) aspect summarisation [9]. This section presents a review of previous work performed in the field of aspect-based OM. As such, the different techniques proposed by researchers are presented.

Hu and Liu [9,20] proposed a novel approach for the extraction and summarisation of frequent aspects. First, part of speech tagging (POS) is performed and then a POS-tagged review is made into an input for the Apriori algorithm for the extraction of frequent aspects. The main purpose of the study was to generate an aspect-based summary of product reviews. However, the approach did not extract adjectives or non-object aspects.

Chinsha et al. [6] proposed an aspect-based OM system using syntactic dependency patterns for aspect extraction and the SentiWordNet (SWN) lexicon for the scoring of aspect-related opinions on a dataset of restaurant reviews. However, due to the incorrect scoring of aspect-related terms by the SWN lexicon, there exists a need to develop a revised scoring mechanism for those opinion words for which the SWN provides an incorrect sentiment score.

A combination of the lexicon- and rule-based techniques was proposed by Ding et al. [9]. In order to calculate the semantic orientation of opinion words, a dictionary of positive and negative sentiment words is used. The negation terms (e.g., no, not, etc.) are handled by applying a set of linguistic rules to shift the polarity of the sentiment words. They used an aggregate scoring scheme to compute the sentiment scores of multiple opinion words pertaining to different aspects.

Maharani et al. [16] proposed a set of syntactic patterns for aspect extraction by manually inspecting features from a review text. They obtained better results in terms of improved precision and recall for extracting aspects. However, they did not address aspect-related sentiment extraction, while an aspect-based summary was also not generated.

In [12], the authors proposed hybrid dependency patterns for aspect extraction from user reviews. Their approach is partially dependent on the syntactic sequence as well as the semantic relation. They used polarity adjectives for the semantic relation and linguistic patterns for the syntactic sequence, thereby achieving better results with respect to the comparing methods alternative methods. However, they used only a limited number of heuristic patterns and they did not perform aspect-related sentiment classification. Moreover, summary generation was also not performed. Our work represents an extension of their work due to proposing an extended set of patterns and assigning sentiment scores to opinions expressed on aspects.

Htay et al. [8] suggested a novel set of pattern-based rules for the extraction of aspects and the associated sentiment words and phrases from product reviews. On comparison with the state-of-the art techniques, their technique improved the precision, recall and f-measure. However, they only worked on aspect extraction, without computing the sentiments expressed with regards to the corresponding aspects.

Samha et al. [27] proposed a lexicon-based approach for the sentiment scoring of aspect-related opinion words using an opinion lexicon and a manually constructed sentiment lexicon. The opinion lexicon they used contained only a limited number of opinion terms, while no polarity scores were attached to the opinion terms. Furthermore, the scores for the opinion terms were assigned manually. More accurate results could be achieved by using a lexicon-based and corpus-based hybrid approach for the sentiment scoring of aspect-related opinion terms.

Qiu et al. [25] proposed a double propagation method for extracting an aspect and the related opinions at the same time. As opinion-related words are used to express sentiments regarding aspects, their technique works by establishing a natural association between opinion-related words and aspects. The input to the system is provided through a set of seed sentiment terms for the extraction of aspects and new sentiment terms. To extract more opinion terms and more aspects, the process is looped back until no more new terms exist. However, as the infrequent aspect extraction is not focused, only low precision and normal recall are achieved.

In [28], a novel feature-based heuristic technique was proposed for the aspect-based sentiment classification of movie reviews. To identify the sentence-level aspect score, the authors proposed an SWN-based algorithm using a pattern combination of Adverb + Adjective + Adverb + Verb. A POS tagger is applied to tag and extract the aspect terms in a sentence, while the sentiment terms are identified by searching up to 5-gram forward or backward of the aspect term. After that, the SWN-based score is computed using a combination of (Adverb) + (Adjective) and (Adverb) + (Verb). They used movie reviews and obtained an accuracy of 78.7% when comparing the results with Alchemy API. The major limitation of the technique was the lack of identification of multiple aspects and sentiments.

Kessler and Nicolov [11] performed an aspect-opinion-based annotation of datasets in the camera and car domains. In order to identify the opinions related to the aspects, they trained a machine learning classifier (SVM). The feature vectors are constructed by considering the syntactic and semantic association between the candidate aspects and their related opinions. However, their study did not discuss how the sentiment expressions are identified.

Zhuang et al. [34] proposed a supervised learning technique for the extraction of aspect-opinion pairs. An annotated dataset is used to learn/train the sentiments by using a combination of dependency and POS paths that connect the aspect-opinion pairs. They used a movie review dataset for the evaluation of their approach. They compared the results with baseline methods and achieved a higher F-measure of 0.529.

Varghese and Jayasree [32] proposed a dependency parsing technique for the classification of sentences into subjective and objective categories. The SWN lexicon is used for the opinion word extraction and classification. The system is trained by using a support vector machine (SVM) on reviews of digital cameras. Their work was focused on a single domain and they only considered explicit aspects. The method achieved an average accuracy of 77.98%. In order to achieve good accuracy results, the approach requires a large amount of training data; otherwise, it may fail.

In [30], a linguistic technique was proposed for aspect-level OM. A dependency tree is used to split sentences into clauses as well as to explore the grammatical dependencies between words/phrases. Then, SWN and domain-specific lexicons are used for the polarity identification and sentiment scoring. However, the automatic aspect extraction was not addressed.

In [17], the authors proposed a learning-based approach for explicit aspect extraction. They used a decision tree and rule learning for pattern generation based on sequence labelling. The generated patterns are used for the identification and extraction of explicit aspects from customer reviews and the opinion lexicon. The experiment results showed better performance than the baseline methods. However, there was a significant increase in the number of generated patterns when compared with previous patterns.

The existing studies concerning aspect-based OM are mainly based on heuristic patterns for aspect-sentiment extraction [8,12,16,22,26,31], supervised or unsupervised aspect-level sentiment classification [19,23,31] and aspect-based summary generation [20,22,27,31]. The aforementioned approaches to aspect-based OM are characterised by different problems, namely (i) a limited number of heuristic patterns for aspect and sentiment extraction, (ii) the dependency of the aspect-level sentiment classification algorithm on either a lexicon (unsupervised) or a supervised classifier with an annotated dataset, and (iii) a less user-friendly aspect-based summary. Therefore, further work is required to address the aforementioned problems by (i) extending the set of heuristic patterns for aspect-sentiment extraction, (ii) proposing a hybrid sentiment classification algorithm with the additional support of addressing modifiers and negations, and (iii) generating a user-friendly aspect-level summary. The problems associated with the existing approaches can be overcome by proposing an efficient aspect-based OM framework using a hybrid scheme of classification with (i) an extended set of linguistic patterns, (ii) a hybrid sentiment classifier with the extended support of intensifiers and negations, and (iii) a user-friendly summary generator.

3 The proposed framework

The proposed framework (Fig. 1) is comprised of four main modules, namely (i) aspect-sentiment extraction (ASE), (ii)

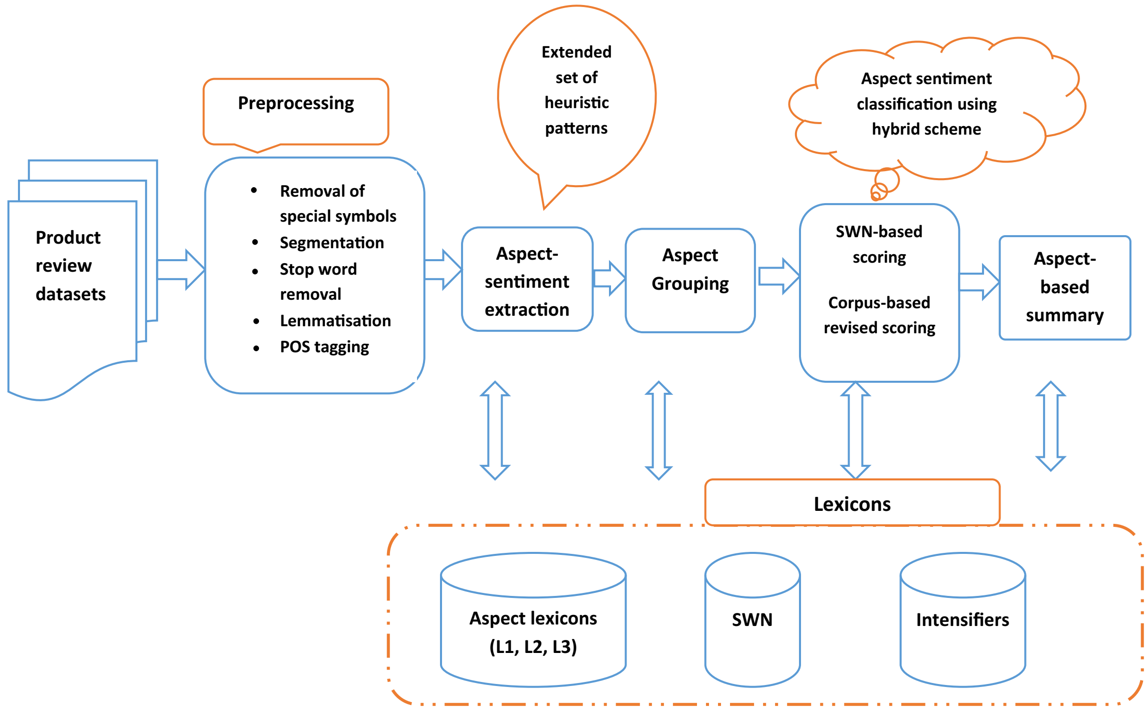


Fig. 1 The Proposed framework

aspect grouping, (iii) aspect-sentiment classification (ASC), and (iv) aspect-based summary generation (ASG).

3.1 Aspect-sentiment extraction

The aspect-sentiment extraction module deals with the detection, extraction and classification of explicit aspects and their associated sentiments. For example, the review sentence "The battery life is good" contains "battery life" as an aspect and "good" as a sentiment word. The proposed aspect-sentiment extraction module represents an enhancement of the work proposed by Khan et al. [12]. They proposed a limited number of patterns. However, as an enhancement of their work, we propose an extended set of heuristic patterns (Table 1), wherein nouns (single word), noun phrases (multi-word) and verbs are considered to be candidate terms for the aspects during their extraction from a given review text.

For example, pattern #8 (P8) shows that if the first word is an adjective (JJ), the second word is "TO" and the third word is a verb (VB), then the third word (i.e. VB) will be an aspect and the first word (JJ) will be an opinion expressed regarding that aspect. For instance, in the sentence "Easy to use", the word "easy" is an adjective (opinion word) and the word "use" is a verb (aspect). Further, P9 shows that if the first word is a verb (VB) and the second word is an adjective (JJ), then the first word (i.e. VB) will be an aspect and the second word (JJ) will be an opinion expressed regarding that aspect. For instance, in the sentence "Looks great", the word "great" is an adjective (opinion word) and the word "looks" is a verb (aspect). Additionally, P10 shows that if the first word is a verb (VB), the second word is a noun (NN) and the third word is also a noun (NN), then the first word (i.e. VB) will be an opinion expressed on an aspect while the second and third words (NN) will be an aspect. For instance, in the sentence "Superb voice quality", the word "superb" is a verb (opinion word) and "voice quality" is the noun phrase (aspect). For each POS-tagged sentence in a given review, it is possible to search and identify the corresponding aspect-sentiment rule listed in Table 1. For example, when the sentence "The pictures are absolutely amazing" is passed through the POS tagger, it tags the words "amazing" as an adjective and "pictures" as a noun. This can be represented as follows: The/DT pictures/NNS are/VBP absolutely/RB amazing/JJ.

In the aforementioned example, the POS tags "pictures/NNS" and "amazing/JJ" match with P3 (Table 1), where the word "amazing" represents an opinion expressed regarding the aspect "pictures". The extracted aspects and their associated opinions in the POS-tagged sentences are stored in an Excel file for further processing, which constitutes our aspect-opinion lexicon (L1). Table 2 presents a partial listing of the L1.

Algorithm 1 is used for the extraction of the aspects and the related sentiments from the aspect-opinion lexicon (L1).

Table 1 Set of extended heuristic patterns for aspect-sentiment detection (aspect-rule-lexicon)

Rule/Pattern#	Patterns			Example (s)
	1st word	2nd word	3rd word	
P1	JJ	NN/NNS	–	
	(Adjective,	Noun)	–	Beautiful pictures, blurry pictures, smooth touch
P2	JJ	NN/NNS	NN/NNS	Simple cell phone
	(Adjective,	Noun,	Noun)	
P3	NN/NNS	JJ	–	Software terrible
	(Noun,	Adjective)	–	
P4	VB	NN/NNS	–	Recommended camera
	(Verb,	Noun)	–	
P5	AVB	JJ	NN/NNS	Very nice pictures
	(Adverb,	Adjective,	Noun)	
P6	NN	NN/NNS	JJ	Battery life lasts
	(Noun,	Noun,	Adjective)	
P7	NN	IN	NN	Quality of photo
	(Noun,	IN,	Noun)	
P8	JJ	TO	VB	Easy to use
	(Adjective,	TO,	Verb)	
P9	VB	JJ	–	Looks great
	(Verb,	Adjective)		
P10	VB	VB	NN	Superb voice quality
	(Verb,	Noun,	Noun)	

Algorithm 1 Abstract of the steps taken for aspect-sentiment extraction

```
Input: POS-tagged pre-processed sentence, aspect-rule-lexicon (Table 1)
Output: Aspect-opinion lexicon (L1)
------------------------------------------------------------------------------------------------
Begin
1     for each POS-tagged-sentence "si" in POS-tagged sentences do
        begin
2          while not (aspect-rule-lexicon.eof)
             begin
3                 if (POS-tagged-sentence "si" matches rule in aspect-rule-lexicon) then
                     begin
4               Extract aspect-sentiment N-gram along with rule # from aspect-rule-lexicon
5               Store it in Aspect-Opinion-Lexicon (L1)
                    End if
           End while
      End for
End
```

While performing different experiments on the datasets, we identified heuristic patterns for the extraction of aspects and the related sentiments on the basis of the observations made. As we conducted our experiments on benchmark datasets, the proposed heuristic patterns presented in Table 2 work well (see results in Table 13) for the current study with the abovementioned datasets. However, more patterns could be derived if we changed the datasets. For this purpose, a more generalised framework could be devised as a future work, wherein the user could be given the option to add new

Table 2 Examples of review sentences with the matching patterns (aspect-opinion lexicon [L1])

S#	POS-tagged sentence	Aspect(s)	Opinion	Patterns
1	The/DT pictures/NNS are/VBP absolutely/RB amazing/JJ.	Pictures	Amazing	P3
2	Long/JJ battery/NN life/NN	Battery life	Long	P2
3	Ear/NN phones/NNS are/VBP very/RB comfortable/JJ	Ear phone	Comfortable	P6
4	Software/NN is/VBZ very/RB terrible/JJ	Software	Terrible	P3
5	Its/PRP light/JJ weight/NN enough/RB to/TO carry/VB easily/RB	Weight	Light	P1
6	The/DT battery/NN life/NN is/VBZ good/JJ	Battery life	Good	P6
7	The/DT pictures/NNS are/VBP very/RB vibrant/JJ	Pictures	Vibrant	P3
8	The/DT phone/NN has/VBZ superb/VB voice/NN quality/NN	Voice quality	Superb	P10
9	The/DT camera/NN is/VBZ easy/JJ to/TO use/VB	Use	Easy	P8
10	The/DT phone/NN looks/VB great/ JJ	Looks	Great	P9

rules for aspect-sentiment extraction on the basis of patterns generated dynamically from the given dataset.

3.2 Aspect grouping

After the extraction of the aspects, the next step is to group similar aspects. Usually, users express their opinions regarding the same aspect by using different words or phrases. For instance, in the mobile phone dataset, "battery", "battery life" and "battery usage" represent different expressions that refer to the same aspect, that is, "battery". In order to produce an effective aspect-based opinion summary, such words (synonyms) concerning the aspects should be grouped [33]. The proposed aspect grouping technique is based on the semantic similarity measure [13]. To group similar aspects in the aspect lexicon, we iterate through the intermediate lexicon: "aspect-opinion lexicon". We compare each of the input aspect words (aw_i) with the next aspect word in the aspect lexicon by computing the semantic similarity using the co-reference PMI_score [13].

This can be computed as:

$$coref\_sim\_score\,(w1, w2) = \frac{S^{\beta}\,(w1)}{\beta 1} + \frac{S^{\beta}\,(w2)}{\beta 2} \qquad (1)$$

In Eq. 1, w1 and w2 represent the two words between which the semantic similarity is to be computed. $S^{\beta}\,(w1)$ and $S^{\beta}\,(w2)$represent the summation of all the positive PMI scores of the whole collection of semantically similar terms. $\beta 1$ and $\beta 2$ indicate the existence of word w in the text, while the coref_sim_score () function provides a numeric similarity score between the two words. This score ranges from 0 to 1. In this step, we choose words whose coref_sim_score () exceeds 0.5, which is a manually tested threshold (Table 3).

Table 3 Words and their associated synsets based on the co-reference PMI_score

Input word (awi) in the aspect-opinion lexicon	Related n-gram (awi+1…..awi+n) in the aspect-opinion lexicon	$coref\_sim\_score$ (using Eq. 1)
Picture	Picture quality	0.94
	Snap	0.83
	Image	0.64
	Photograph snap-shot	0.52
	Scene	0.48

It is selected after the manual inspection of the input word and its related synonyms. Different values (in the range of 0 to 1) for the threshold are inspected both for the input word (e.g. picture) and the related synonyms' bigrams (e.g. picture quality). Finally, we reach the conclusion that the ideal values of the $coref\_sim\_score$ (Eq. 1) range between {0.5 and 1}. Therefore, words with a semantic similarity score greater than 0.5 are grouped together.

The selected words are grouped together and then placed in the "aspect-synonym" column of Table 4. Furthermore, the corresponding sentence IDs of a particular aspect group are also combined and recorded in the "sentence-group#" column. Each of the sentence group numbers is tagged with a group title, where the "group-title" is the name of the first aspect in the "aspect-synonym" column. For example, the aspect-group title for sentence-group# "S1-S12-S18" is labelled as "Pictures" because "pictures" is the first word in the aspect-group column. Table 4 presents a sample set of entries in the aspect-opinion-group lexicon (L2) containing similar aspects and their grouping.

Table 4 Partial listing of aspect-opinion-group lexicon (L2)

Sentence-group#	POS-tagged sentence	Aspect(s)	Aspect synonyms	Aspect-group-title
S1-S12-S18	The/DT pictures/NNS are/VBP absolutely/RB amazing/JJ.	Pictures	Pictures	Pictures
	The/DT picture/NN quality/NN is/VBZ good/JJ	Picture Quality	Picture quality	
	The/DT snaps/NNS are/VBP superb/JJ	Snaps	Snaps	
S2-S16-S29	Long/JJ battery/NN life/NN	Battery life	Battery life	Battery life
	The/DT battery/NN goes/VBZ down/RB on/IN full/JJ brightness/NN	Battery	Battery	
	The/DT battery/NN power/NN of/IN S4/NNP is/VBZ good/JJ	Battery power	Battery power	

Algorithm 2 Aspect grouping

```
Input: Aspect-opinion lexicon (L1)
Output: Group of similar words for a given aspect word
Begin
1     for each aspect-word "awi " of the aspect-opinion – lexicon (L1) do
2          compute co-reference PMI_score using Eq. 1 between awi and awi+1
3      if (the similarity score computed at step# 2 is greater than 0.5) then
       begin
4             Select the word awi+1, and append it in front of corresponding input aspect-
word (awi) (aspect-group column) of aspect-opinion lexicon.
5    combine related sentence-ids and place them in the "sentence-group#" column.
6             Assign awi to aspect-group-title column
7             Store it in Aspect-Opinion-Group-Lexicon(L2)
           End if
      End for
End
```

Algorithm 2 determines the synonym groups.

3.3 Aspect-sentiment classification

In order to assign sentiment scores to the opinions expressed on the extracted aspects in a given sentence, a hybrid sentiment scoring technique is proposed, which combines lexicon-based and corpus-based concepts. Asghar et al. [3] proposed a hybrid sentiment scoring technique by combining the SWN lexicon with a revised mutual information measure. However, we propose to combine the SWN lexicon with the probability-driven revised scoring technique [1]. The hybrid sentiment scoring technique operates as follows. First, the sentiment scores of the aspect-related opinion terms are computed using the SentiWordNet lexicon. If the SWN lexicon yields incorrect sentiment scores, then the probability-driven corpus-based sentiment scoring technique is used to assign sentiment scores to the aspect-related opinion terms. This module works as follows.

3.3.1 SentiWordNet (SWN)-based scoring

The SWN-based scoring technique uses the SentiWordNet lexicon to assign polarity scores to the opinion words expressed regarding product aspects. For example, the sentence "The camera produces a blurred picture" contains the negative opinion word "blurred", which is expressed about the aspect "picture" of the product "camera". The SWN lexicon is chosen due to its wide coverage of words and their sentiment scores. It is a general purpose lexicon containing more than 60,000 words as well as being publically available [7].

In the SWN lexicon, a given word can have multiple senses. To evaluate the correct sense of an opinion word with multiple senses, we take three polarity scores, that is, positive, negative and objective scores of all senses of all the POS-tagged words available in the SWN lexicon. We compute three average values: Pos_{Score} , Neg_{Score} and Obj_{Score} for all the synsets of word "w_i" with respect to a particular part-of-speech (POS).

$$Sum_{Pp}\,(w_i) = \sum_{i=0}^{n} P\,(i) \tag{2}$$

$$Sum_{Np}\,(w_i) = \sum_{i=0}^{n} N\,(i) \tag{3}$$

$$Sum_{Op}\,(w_i) = \sum_{i=0}^{n} O\,(i) \tag{4}$$

Table 5 Example sentence with POS tagging, aspect and opinion

Example sentence	Long battery life.
POS-tagged sentence	Long/JJ battery/NN life/NN
Aspect	Battery life
Opinion	Long

$$Pos_{Score}\left(w_i\right) = \frac{Sum_{Pp}\left(w_i\right)}{totalSyn\left(w_i\right)} \tag{5}$$

$$Neg_{Score}\left(w_i\right) = \frac{Sum_{Np}\left(w_i\right)}{totalSyn\left(w_i\right)} \tag{6}$$

$$Obj_{Score}\left(w_i\right) = \frac{Sum_{Op}\left(w_i\right)}{totalSyn\left(w_i\right)} \tag{7}$$

where $Sum_{Pp}\left(w_i\right)$, $Sum_{Np}\left(w_i\right)$ and $Sum_{Op}\left(w_i\right)$ are the cumulative polarity scores of the ith positive, negative and objective synsets of word w_i, while $totalSyn$ is the total number of synsets of a particular POS of the term w_i. We take the following example (Table 5).

Table 5 shows that the POS tag for the opinion term (long) is an adjective. The SWN lexicon has twelve entries for the word "long", one sense for the verb category, two for the adverb category and nine senses for the adjective category. Using Eqs. 5, 6 and Eq. 7, the average positive, negative and objective scores can be computed as follows.

$$\begin{aligned} Pos_{Score}\,(long) &= \frac{0.125{+}0{+}0 + 0{+}0 + 0.250{+}0.375 + 0.00 + 0.250}{9} \\ &= 0.111 \end{aligned}$$

$$Neg_{Score}\,(long) = \frac{0.375 + 0 + 0 + 0 + 0 + 0 + 0.5 + 0.5 + 0.12}{9} = 0.167$$

$$\begin{aligned} Obj_{Score}\,(long) &= \frac{0.5 + 1 + 1 + 1 + 1 + 0.75 + 0.875 + 0.5 + 0.63}{9} \\ &= 0.722 \end{aligned}$$

Thus, the average positive, negative and objective scores for all senses of the word "long" with a POS verb are:0.111, 0.167 and 0.722, respectively.

To compute the final polarity score for a given word, we choose the dominant polarity score of word "w_i" as:

$$SWN_{Score}\left(w_i\right) = \begin{cases} Pos_{Score} & if \max\left(Pos_{Score}, Neg_{Score}, Obj_{Score}\right) = Pos_{Score} \\ Neg_{Score} & if \max\left(Pos_{Score}, Neg_{Score}, Obj_{Score}\right) = Neg_{Score} \\ Obj_{Score} & otherwise \end{cases} \tag{8}$$

The "SW$N_{\text{Score}}\left(w_i\right)''$ is positive if the average positive score $\left(Pos_{Score}\right)$ is greater than both the average negative $\left(Neg_{Score}\right)$ and objective $\left(Obj_{Score}\right)$ scores, otherwise it is negative. The polarity score is considered to be$objective(Obj_{Score})$ if the average positive and negative polarity scores are equal or the objective polarity score is greater than the positive and negative scores. As in the above example, the polarity scores $(Pos_{Score}, Neg_{Score}, Obj_{Score})$ for the term "blurred" are 0.111, 0.167 and 0.722, respectively. Therefore, the SWN_{Score} (long) =0.722, which is objective.

Yet, there exists a problem. The SWN-based dominant score of the opinion word "long" is objective, which is incorrect, since it expresses a positive sentiment about the aspect "Battery life". Therefore, a revised scoring mechanism is required to assign accurate sentiment scores to such opinion words. Table 6 shows that among the SWN-based scores listed in the last column, five are correct and five are incorrect. For example, the SWN-based scores for the opinion words "amazing", "terrible", "good", "vibrant" and "superb" are correctly computed, while the rest of the opinion words are assigned incorrect polarity scores. Hence, a revised scoring technique is required to assign accurate scores to the opinion words expressed regarding product aspects. The detail of the revised scoring technique is presented in the next section (Sect. 3.3.2).

3.3.2 Revised scoring scheme

The revised scoring scheme aims to assign accurate sentiment scores to those opinion words for which the SWN lexicon yields incorrect sentiment scores. For this purpose, we propose a corpus-based revised scoring scheme, which operates in two steps: (i) detection of the sentiment class of opinion words, and (ii) sentiment scoring of opinion words.

Sentiment class detection of opinion words: The majority of words in a particular domain have one polarity class in the SWN lexicon, whereas their occurrence in a labelled corpus shows a strong tendency with the other sentiment classes. For example, if a given term's SWN-based average dominant polarity score is negative, but its occurrence in the positive documents is greater when compared to the negative ones, we change its polarity score. In order to check which words exhibit a higher frequency in a particular sentiment class as compared to the other classes, we calculate the frequency-driven probability [1] between each term in the test reviews, and its sentiment class is detected in the training documents as follows:

$$sentiment\,(ow) = \begin{cases} +ive, if\, Prob\,(ow, c_+) > Prob(ow, c_-) \\ -ive, otherwise \end{cases} \tag{9}$$

Table 6 SWN-based sentiment scoring of aspect-related opinion words

S#	POS-tagged sentence	Aspect(s)	Opinion	SWN-based score (w_i)
1	The/DT pictures/NNS are/VBP absolutely/RB amazing/JJ	Pictures	Amazing	Positive (0.688)
2	Long/JJ battery/NN life/NN	Battery life	Long	Objective (0.722)
3	Ear/NN phones/NNS are/VBP very/RB comfortable/JJ	Ear phone	Comfortable	Objective (0.45)
4	Software/NN is/VBZ very/RB terrible/JJ	Software	Terrible	Negative (0.656)
5	Its/PRP light/JJ weight/NN enough/RB to/TO carry/VB easily/RB	Weight	Light	Objective (0.65)
6	The/DT battery/NN life/NN is/VBZ good/JJ	Battery life	Good	Positive (0.619)
7	The/DT pictures/NNS are/VBP very/RB vibrant/JJ	Pictures	Vibrant	Positive (0.625)
8	The/DT phone/NN has/VBZ superb/VB voice/NN quality/NN	Voice quality	Superb	Positive (0.875)
9	The/DT camera/NN is/VBZ easy/JJ to/TO use/VB	Use	Easy	Objective (0.458)
10	The/DT phone/NN looks/VB great/ JJ	Looks	Great	Objective (0.667)

where ow is an opinion word in the test corpus and Prob(ow, c_+) and Prob(ow, c_-) are the probabilities of word ow occurring in positive and negative documents of the training set, respectively.

$$Prob\,(ow, c_+) = \frac{freq\,(ow \in D_+)}{|D_+|} \tag{10}$$

$$Prob\,(ow, c_-) = \frac{freq\,(ow \in D_-)}{|D_-|} \tag{11}$$

D_+ and D_- are the training set of +ive and -ive documents, repectively.

Sentiment scoring of opinion words: The sentiment scoring module aims to assign sentiment scores to those opinion words whose opinion score is incorrectly given in the SWN lexicon. In order to assign the correct sentiment score to an opinion word, we alter the sentiment score of that opinion word to its opposite pole based on its occurrences in the positive and negative documents. For example, if an opinion word's SWN-based polarity (Eq. 8) is negative, but it has a higher frequency in positive reviews (Eq. 9), then we change it to a positive sentiment and vice versa. Furthermore, if a given word's SWN-based sentiment is objective (Eq. 8), but it has a higher frequency in negative documents (Eq. 9), then the SWN-based objective score is shifted to its respective positive score, which is also the case for words in negative documents. Its score is computed as follows:

Let a=Pos_{Score}
b=Neg_{Score}
c=Obj_{Score},
then domain specific sentiment score is computed as follows:

$$sentiment(ow)^{DS} = \begin{cases} Pos_{Score}, & (a > b) \wedge (a > c) \wedge \Pr ob\,(ow, c_+) > \Pr ob\,(ow, c_-)) \\ Pos_{Score} \times (-1), & (a > b) \wedge (a > c) \wedge ((\Pr ob\,(ow, c_-) > \Pr ob\,(ow, c_+)) \\ Neg_{Score} \times (-1), & (b > a) \wedge (b > c) \wedge ((\Pr ob\,(ow, c_+) > \Pr ob\,(ow, c_-)) \\ Neg_{Score}, & (b > a) \wedge (b > c) \wedge ((\Pr ob\,(ow, c_-) > \Pr ob\,(ow, c_+)) \end{cases} \tag{12}$$

For example, the SWN-based score for "long" is objective ($Obj_{Score}\,(long)$=0.722), which is incorrect. As its frequency in the negative documents is higher than in the positive reviews (using Eq. 9), we update its sentiment score (using Eq. 12) to −0.722, which depicts a correct negative polarity score. Table 7 presents the domain-specific scores of those entries for which the SWN lexicon provides an incorrect scoring.

3.3.3 Intensifier handling

Intensifiers are the terms that increase or decrease the intensity of the opinion words in a given sentence. For

Table 7 SWN- and corpus-based scoring of aspect-related opinion words (aspect-opinion-scoring lexicon [L3])

#	POS-tagged sentence	Aspect(s)	Opinion words	SWN-based score (w) using Eq. 8	Corpus-based score using Eq. 12
1	The/DT pictures/NNS are/VBP absolutely/RB amazing/JJ	Pictures	Amazing	Positive (0.688)	–
2	Long/JJ battery/NN life/NN	Battery life	Long	Objective (0.722)	Positive (0.722)
3	Ear/NN phones/NNS are/VBP very/RB comfortable/JJ	Ear phones	Comfortable	Objective (0.45)	Positive (0.45)
4	Software/NN is/VBZ very/RB terrible/JJ	Software	Terrible	Negative (0.656)	–
5	Its/PRP light/JJ weight/NN enough/RB to/TO carry/VB easily/RB	Weight	Light	Objective (0.65)	Positive (0.65)
6	The/DT battery/NN life/NN is/VBZ good/JJ	Battery life	Good	Positive (0.619)	–
7	The/DT pictures/NNS are/VBP very/RB vibrant/JJ	Pictures	Vibrant	Positive (0.625)	–
8	The/DT phone/NN has/VBZ superb/VB voice/NN quality/NN	Voice quality	Superb	Positive (0.875)	–
9	The/DT camera/NN is/VBZ easy/JJ to/TO use/VB	Use	Easy	Objective (0.458)	Positive (0.458)
10	The/DT phone/NN looks/VB great/ JJ	Looks	Great	Objective (0.667)	Positive (0.667)

example, “very”, “somewhat”, “slightly”, “too”, “really” and “extremely” increase or decrease the semantic orientation of the opinion word.

In the present study, we used the 50 English intensifiers proposed by Benzinger [14]. We assigned a polarity score to each intensifier by using the numeric values (e.g. 1, −1, 0.5, −0.5) proposed by [10,29] in order to compile a list of positive and negative intensifiers (Table 8).

Let I_{pos_neg} be a list of positive and negative intensifiers represented as:

I_{pos_neg}= {list of positive and negative intensifiers}

If a given term is present in a list of positive or negative intensifiers, then the sentiment score of the neighbouring opinion word is calculated as follows:

$$\begin{aligned} & sentiment_{score-intensifier}\,(ow) \\ & \quad = sentiment_{score}\,(ow) + sentiment_{score}\,(ow) \\ & \quad \times sentiment\_score\,(w_x)\,, if\,(w_x \in I_{pos\_neg}) \end{aligned} \tag{13}$$

where w_x represents a word belonging to a list of positive and negative intensifiers, ow is an opinion word and sentiment_score(w_x) is a numeric score for the intensifier obtained from the enhancer list I_{pos_neg}. The sentiment score for the neighbouring opinion word is computed by multiplying the numeric score for the intensifier by the SWN-based sentiment score (using Eq. 8) or domain-specific score (using Eq. 12) of an opinion word and then adding it to the SWN-based sentiment score or corpus-based score of an opinion word.

Table 8 Partial list of positive and negative intensifiers

Intensifier	Score	Intensifier	Score	Intensifier	Score
Too	+0.6	Totally	+0.7	Extremely	+0.8
Pretty	+0.3	Less	−0.5	Very	+0.5
Quite	−0.2	Hardly	−1	Slight	−0.4
Completely	+1	Really	+1.5		

For example, in the sentence “Ear phones are very comfortable”, the intensifier “very” increases the sentiment strength of the neighbouring opinion word “comfortable”. Therefore, using Eq. 13, the revised polarity score for the opinion word “comfortable” is computed as follows:

$$\begin{aligned} & sentiment_{score-intensifier}\,(\text{“}verycomfortable''\text{)} = \\ & sentiment_{score}(\text{“}comfortable''\text{)} + sentiment_{score}(\text{“comfortable''}) \times \\ & sentiment\_score(\text{“}very''\text{)} = 0.45 + 0.45 \times 0.5 \\ & \quad = 0.45 + 0.225 = 0.675 \end{aligned}$$

where 0.45 is the sentiment score for the opinion word “comfortable” as retrieved from the SWN lexicon (Eq. 8),

Table 9 Sentiment scoring of aspect-related opinion words with intensifiers and negation terms

S#	POS-tagged sentence	Aspect(s)	Opinion words	Intensifier	Negation	SWN-based score (w) using Eq. 8	Corpus-based score using Eq. 12	Revised sentiment score after applying intensifiers (Eq. 13) and negations (Eq. 14)
1	The/DT Internet/NN speed/NN is/VBZ really/RB good/JJ	Internet speed	Good	Really		Positive (0.619)		Positive (1.55)
2	The/DT battery/NN life/NN is/VBZ not/RB good/JJ	Battery life	Good		Not	Positive (0.619)	–	Negative (0.619)
3	Ear/NN phones/NNS are/VBP very/RB comfortable/JJ	Ear phones	Comfortable	Very		Objective (0.45)	Positive (0.45)	Positive (0.675)
4	Voice/NNP quality/NN is/VBZ too/RB good/JJ	Voice quality	Good	Too		Positive (0.619)	–	Positive (0.99)
5	Screen/NNP resolution/NN of/IN this/DT mobile/NN is/VBZ quite/RB low/JJ	Screen resolution	Low	Quite		Objective (0.612)	Negative (0.612)	Negative (0.489)

0.5 is the weight of the positive intensifier "very" as obtained from Table 8, and 0.675 is the revised score for the opinion word "comfortable" after applying the intensifier.

3.3.4 Negation management

When expressing sentiments about the aspects of a given product, users often make use of negation terms, including "not", "no", "cannot", "could not" and "never", which shift the polarity of the opinion word to the opposite pole. For example, the sentences "the battery life is good" and "the battery life is not good" have different semantic orientations. The first sentence bears a positive polarity; however, in the second sentence, the negation term "not" switches the polarity of the sentiment word "good" from positive to negative. Therefore, negation terms need special attention if an accurate sentiment classification is to be achieved. Similar to the work performed by [4] in relation to negation handling, a list of negation words is created and the presence of such words in a sentence is checked.

Let Ng_list be a list of negation terms defined as:

Ng_list ={List of negation terms}

If a given term is present in the negation list, then the polarity of the neighbouring sentiment word is switched to the opposite pole by multiplying the score of the sentiment word by −1 as follows:

$$sentiment_{score-neg}\,(ow) = (sentiment_{score}\,(ow) \times (-1)),\quad if((ow-1) \in Ng\_list) \tag{14}$$

where ow denotes the neighbouring opinion word and ow-1 denotes the predecessor word of an opinion word belonging to a list of negation terms Ng_list. For example, using Eq. 14, the polarity score for "not good" is computed as follows:

$$\begin{aligned} &\mathrm{sentiment}_{\mathrm{score-neg}}\left(\text{“not good''}\right)\\ &= \left((\mathrm{sentiment}_{\mathrm{score}}\left(\text{“good''}\right) \times (-1)\right)\\ &= \quad 0.619\mathrm{x} - 1 = -0.619. \end{aligned}$$

Table 9 demonstrates the impact of intensifiers and negation terms on the polarity scores.

Algorithm 3 is used for the classification of the aspect-related sentiments.

Algorithm 3 Aspect-sentiment classification

```
Input: Aspect-opinion lexicon (L2)
Output: Aspect-sentiment classification
Words description: PP = positive polarity, NP = negative polarity, NUP = neutral polarity
-----------------------------------------------------------------------------------------
Function Aspect-sentiment classification (L2)
    For each word in L2 do
      begin
          Search word in SentiWordNet
          Using Eq. 2 compute sum of PP for all senses for particular parts-of-speech of the word
          Using Eq. 3 Using Eq. 3 compute sum of NP for all senses for particular parts-of-speech of the word
          Using Eq. 4  compute sum of NUP for all senses for particular parts-of-speech of the word
          Using Eq. 5 compute Avg-PP
          Using Eq. 6 compute Avg-NP ,
          Using Eq. 7 compute Avg-NUP
//        Apply  Eq. 8
          If (Avg-PP is dominant )Then
                  Word-orientation= "positive"
          Elseif (Avg-NP is dominant) Then
                  Word-orientation= "negative"
          Else
                  Word-orientation= "objective"
          Next word
          If (SWN-based score of word w is incorrect )  Then
              begin
                  Detect new polarity class using Eq. 9
                  Perform corpus-based scoring using Eq. 12
          End if
          Perform intensifiers handling using Eq. 13
          Perform negation handling using Eq. 14
          Store it in Aspect-Opinion-Scoring-Lexicon (L3)
    End for
End Function
```

3.4 Aspect-based summary generation

Summarisation is the final step in the aspect-based opinion mining process [27]. Opinions are summarised on the basis of a particular aspect. A summary is created based on the outcomes of the previous tasks in which the extracted aspects and their corresponding opinions along with their sentiment scores are accumulated. The summary could be visualised in various formats, for example, bar charts, text, or graphs. The proposed output summary will take the form of merits and demerits, where the merits indicate the set of positive product aspects and the demerits represent the set of negative aspects along with their polarity score. The summary generation module (algorithm 4) works as follows. (i) The user enters a query for a particular product for which an aspect-based summary is to be displayed. (ii) For the particular product entered by the user during step #1, the aspect-opinion scoring lexicon (L3) is searched for all the aspects and their group synonyms along with their sentiment scores and polarity classes. (iii) The sentiment scores of each aspect group partnering to a product specified by the user are accumulated and the total sentiment score of an aspect is obtained. (iv) If the total score is positive, then it is considered to be a pro/merit, otherwise it is considered to be a con/demerit. The partial list of such aspects along with their sentiment scores and polarity classes for the user input query "Samsung Galaxy S3" is reported in Table 10. Finally, (iv) the aggregated sentiment scores of the aspects are transformed into user-friendly descriptions with the following scale: Average (0.1–0.49), Good (0.5–1.0), Excellent (>1.0), Not good (−0.1 to −0.49), Poor (−0.5 to −0.99), Bad (> −1.0).

Table 10 Partial list of aspects along with their sentiment scores and polarity classes

Sentence-group#	POS-tagged sentence	Aspect(s)	Sentiment score	Polarity	Aspect group	Aspect-group-title
S1-S7-S10	The/DT camera/NN produces/VB blur/ADJ pictures/NN	Pictures	-0.729	Negative	Pictures	Pictures
	The/DT picture/NN quality/NN is/VBZ good/JJ	Picture quality	0.619	Positive	Picture quality	
	The/DT snaps/NNS are/VBP superb/JJ	Snaps	0.875	Positive	Snaps	
	Total score		0.765	Positive		

A sample aspect-based summary generated using our proposed system is shown in the final part of Fig. 2

The main working algorithm (Algorithm 5) operates by calling all the modules of the proposed framework.

Algorithm 4 Aspect-based product summary generation.

```
Input: Aspect-opinion lexicon (L3)
Output: Aspect-based product summary (Summarytable)
------------------------------------------------------------------------------------------
Begin
    while not (L3.eof ) do
    begin

        For each (sentence-group# in L3)    do
        begin
            add("Aspect-group-title") to Summarytable
            Retrieve sentiment-score of each aspect from L3
            Aggregate sentiment-score of all related aspect under the column "polarity"
            If (aggregate-score of aspect is in the range of {0.1 to 0.49})  Then
                add Merit ←"Average" to Summarytable.
            Elseif(aggregate-score of aspect is in the range of {0.5 to 1.0}) Then
                add.Merit ←"Good" to Summarytable
            Else
                add.Merit ←"Best" to Summarytable
            EndIf
            If (aggregate-score of aspect is in the range of {-0.1 to -0.49})  Then
                add Demerit ←"Not Good" to Summarytable.
            Elseif(aggregate-score of aspect is in the range of {-0.5 to -1.0}) Then
                add.Demerit ←"Poor" to Summarytable

            Else
                add.Demerit ←"Bad" to Summarytable
        End if
    End For
                Display contents of Summarytable.
    End while
End
```

Algorithm 5 Aspect-based opinion mining using a hybrid classification scheme

```
Input: Review sentences
Output: Aspect-based product summary
--------------------------------------------------------------------------------

Main Procedure()
Begin
        Call "Preprocessing(Review Sentences)" module
        Call "Aspect-Sentiment Extraction (Preprocessed Review Sentences)" module
        Call "Synonyms-based Aspect Grouping (Aspect-Sentiment Pair )" module
        Call "Aspect-Sentiment Classification(Synonym-based grouped Aspect-Sentiment
Pair)" module
        Call "Aspect-based Product Summary(Classified Aspect-Sentiment Pair)" module
End
End  Main
```

4 Evaluation and results

This section discusses the experimental setup, datasets and statistical parameters utilised in the present study.

4.1 Experimental setup

Experiments are performed on three product review datasets in order to evaluate the performance of the proposed method. For the implementation, we use the NLTK (Natural Language Toolkit version 3.0) and Python language [5]. To conduct our experiments, we utilise N-fold cross-validation by means of splitting the datasets into M-folds, where one fold is used for testing and the remainder of the M-1 are used for training the system. We use 10-fold cross-validation, where 10% (one fold) is used for testing and 90% (9 of the 10 folds) for training. Finally, an average of all experiments is obtained.

4.2 Datasets and preprocessing

We use three publically available datasets, namely Mobile Phone, MP3 Player and Digital Camera, to conduct the experiments and evaluate the effectiveness of the proposed methods [9]. Each dataset consists of approximately 260 opinion-related sentences written by 325 customers. The unstructured text files containing these datasets are arranged as positive, negative and aspect-wise review sentences along with their polarity scores. In the datasets, the review sentences concerning the product are labelled with its identified aspects/features, along with the positive or negative opinion score (i.e. rating ranges from +3 [strongest] to −3 [weakest]). Then, after ## sign the review sentence starts. Here, in the above sample review sentences, "software" represents the aspect of the MP3 Player and its positive polarity is mentioned before the actual review sentence. During the preprocessing module, in addition to the other steps, the # signs and numeric scores are detected and removed in order to render the dataset useable for our experiments. The sample listing of the aforementioned datasets is presented in Table 11.

User reviews include a considerable amount of noisy data, which have to be preprocessed in order to extract useful information, such as the aspects and opinions, from the reviews. Preprocessing aims to clean the text from a syntactical point of view so that the original sentence structure is not altered [2]. It therefore assists in improving the accuracy of the aspect-based opinion mining process. The preprocessing steps applied in this work include stop words removal, sentence and words splitting, lemmatisation and parts of speech tagging. We use the Python-based NLTK toolkit [5] to implement the preprocessing module.

4.3 Statistical parameters

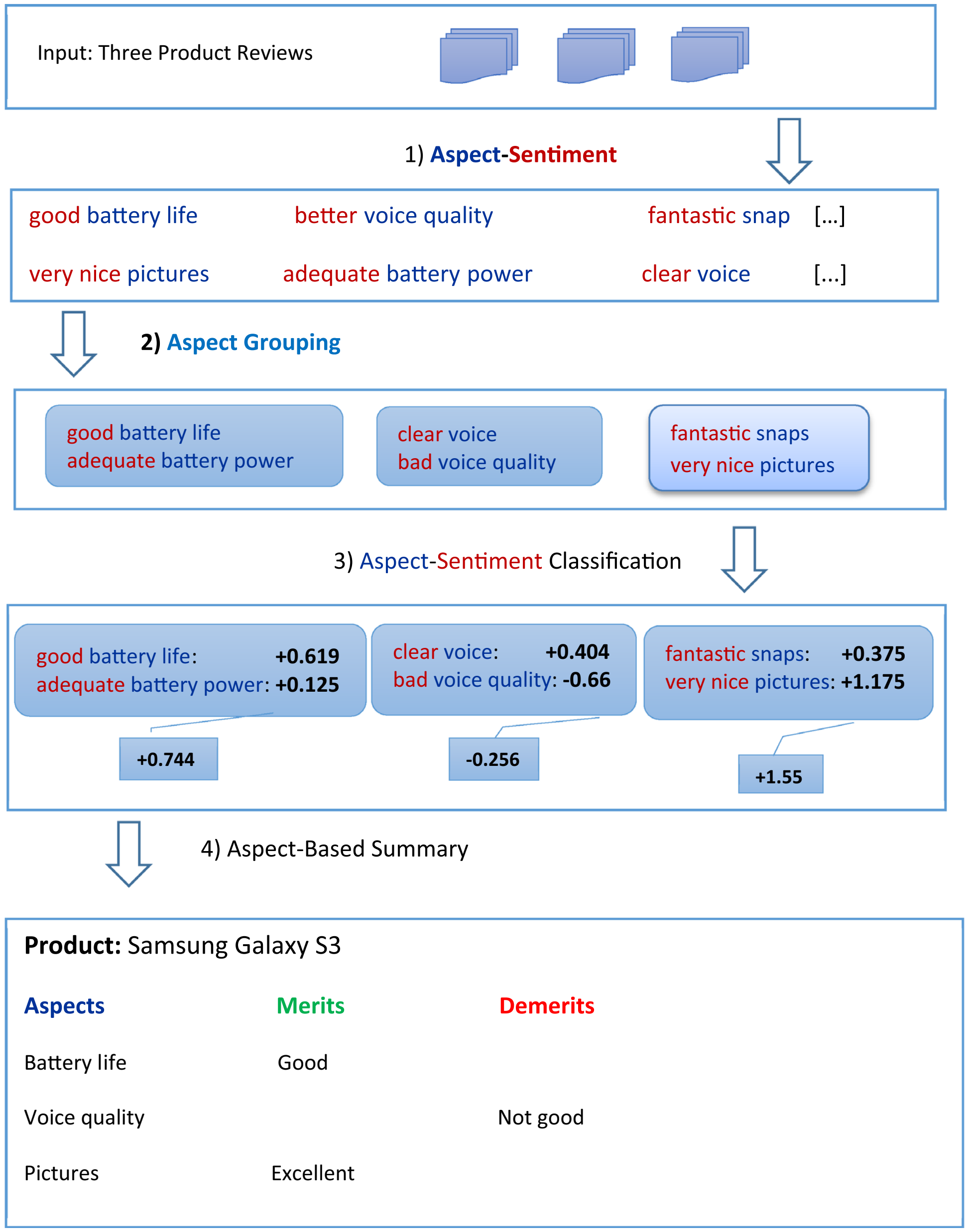
To evaluate the performance of the proposed system, three well-known metrics are used, namely (i) precision, (iii) recall and (iii) F1-measure [10].

4.3.1 Precision

Precision represents the rate at which a classifier makes correct predictions [22]. For example, for a positive class, a high level of precision means more true positive cases and less false positive cases. Using confusion matrix notations, the precision for a negative class is computed as:

$$\text{Precision} = \frac{\text{Tp}}{\text{Tp} + \text{Fp}} \tag{15}$$

Fig. 2 Aspect-based OM and summarisation



or

$$\text{Precision} = \frac{|\text{Extracted\_Aspect} \cap \text{True\_Aspect}|}{|\text{Extracted\_Aspect}|} \tag{16}$$

4.3.2 Recall

Recall is defined as the ratio of instances predicted by a classifier in relation to one class over the total available instances of that class [24]. For a negative class, it is computed as follows:

$$\text{Recall} = \frac{\text{Tp}}{\text{Tp} + \text{fn}} \tag{17}$$

where Tp and fn are the notations described in the "confusion matrix" section (Table 12), or

$$\text{Recall} = \frac{|\text{Extracted\_Aspect} \cap \text{True\_Aspect}|}{|\text{True\_Aspect}|} \tag{18}$$

4.3.3 F1-measure

The F1-measure integrates the precision and recall information. If the precision and recall of a classifier are high, then the F1-measure must be high and vice versa [27]. Mathematically, the F1-measure can be calculated as:

$$\text{F1} - \text{measure} = \frac{2\,(\text{p})\,(\text{r})}{\text{p} + \text{r}} \tag{19}$$

Table 11 Sample review sentences of Hu and Liu (2004)

Products	Review text
MP3 Player	software[+2]##software is easy to use (although redhat software is better, but costs money)
Mobile Phone	backlight[-1]##the backlight on the phone goes off way too quickly (dangerous when you 're driving at night), and there's no way to change this –
Digital Camera	camera[+3]##best camera ever. From the image quality, colour, function, I can say almost everything.

Table 12 Confusion matrix

Predicted value	Actual value Positive	Negative
Positive	Tp	Fn
Negative	Fp	tn

The confusion matrix (Table 12) provides information about correctly and incorrectly classified/predicted cases in order to visualise the performance of an algorithm [4].

In the confusion matrix, Tp, fp, tn and Fn are described as follows: (i) True positives (Tp): instances belonging to the positive class and labelled as positive by the system, (ii) False positives (fp): instances belonging to the negative class and labelled as positive by the system, (iii) True negatives (tn): instances belonging to the negative class and labelled as negative by the system, (iv) False negative (Fn): instances belonging to the positive class and labelled as negative by the system.

4.4 Results and evaluation

We compare our proposed method with the following baseline methods for aspect extraction.

(i) Syntactic Patterns: Warih et al. [16] extracted the explicit aspects from reviews using 11 patterns, which work on the current part of speech tag, (current-1) part of speech tag and (current + 1) part of speech tag. Of the proposed 11 patterns, only seven produced promising results in terms of accurate explicit pattern extraction.

(ii) Pattern Knowledge: Haty et al. [8] worked on the extraction of product aspects and their associated opinions using the pattern knowledge of eight patterns, for example, nouns and noun phrases for aspects and nearest adjective, adverbs, and verbs for opinions. They achieved an average precision of 0.73 for a dataset containing five products.

(iii) Hybrid Dependency Patterns: According to this method, six hybrid dependency patterns are proposed for the extraction of product features by exploiting the lexical relations and opinion context [12].

(iv) Extended Heuristic Patterns (proposed): As shown in Table 13, the proposed method achieved more robust and more accurate results by proposing an extended set of heuristic patterns for the extraction and classification of aspects and their related sentiments.

We also compare our method with the following baseline methods for aspect-related sentiment classification.

(i) Lexicon-based (SentiWordNet): Chinsha et al. [6] used the SentiWordNet lexicon to score the aspect-related opinions in a dataset of restaurant reviews. Incorrect scores were generated for certain aspect-related aspect terms.

(ii) Lexicon-based (Opinion Lexicon): Samha et al. [27] proposed a lexicon-based approach for the sentiment scoring of opinion words using an opinion lexicon and a manually constructed sentiment lexicon. The opinion lexicon they used contains only a limited number of opinion terms and no polarity scores are attached to those opinion terms. Furthermore, the scores for the opinion terms are manually assigned.

(iii) Hybrid (Lexicon-based + Corpus-based): The aforementioned lexicon-based approaches yield incorrect sentiment scores for the aspect-related opinions in certain cases. Therefore, a corpus-based revised sentiment scoring technique is used to assign accurate scores to the expressed aspect-related opinion words. The results obtained are more efficient when compared to the comparing methods (Table 14).

The following deficiencies are identified in the obtained results.

(1) Implicit aspect extraction: The proposed technique for aspect extraction provides only the explicit aspects of matched sentences for processing. The implicit aspects are not identified. For example, "The laptop is heavy to carry". Here, the implicit aspect "weight" cannot be determined using the proposed system.

(2) Multiple aspect sentiments: Our system lacks the ability to address multiple aspects and the related sentiments that are present in a single sentence. For example, in the sentence "The touchscreen is smooth but the battery life is not good", "touchscreen" and "battery life" are the two aspects, while "smooth" and "not good" are the

Table 13 Comparison of aspect extraction (proposed technique) with baseline methods

Study	Technique	Dataset	Avg. P (%)	Avg. R (%)	Avg. F (%)
Htay et al. [8]	Pattern Knowledge (Bag of words + trigram)	MP3 Player	73.3	85.7	79.0
	No. of patterns = 8	Camera			
		Mobile Phone			
Warih et al. [16]	Syntactic Patterns (Bag of words + trigram)	Camera MP3 Player	62.6	72.8	67.2
Khan et al. [12]	Hybrid Dependency Patterns (Dependency relation)	MP3 Player Camera Mobile Phone	78.98	71.77	75.19
Proposed (extended set of heuristic patterns)	Heuristic Patterns (Bigram + trigram)	MP3 Player Mobile Phone Camera	83.46	71.0	77.16

Table 14 Comparison of the aspect-related sentiment classification (proposed technique) with the baseline methods (P: Precision, R: Recall)

Study	Technique	P (%)	R (%)	F (%)
Chinsha et al. [6]	Lexicon-based	79.75	90.1	0.78
Samha et al. [27]	Lexicon-based	0.56	0.61	0.60
Proposed work	Lexicon-based + Corpus-based (Hybrid) with the additional support of intensifiers and negation handling	0.85	0.73	0.78

opinions expressed regarding those two aspects, respectively.

(3) Informal opinion carriers: Informal opinion carriers such as emoticons and slang are not addressed. I incorporated, such informal opinion carriers may serve to improve the efficiency of the aspect-related sentiment classification. For example, the sentence “Wow! Love Dell laptop:) Enjoying its touchscreen”, contains “Wow” as slang and “:)” as an emoticon, which require proper classification for more accurate results.

(4) Redundant aspect candidate terms: Due to the lack of a proper post-processing module, several redundant terms for the aspects are generated by the proposed aspect-sentiment algorithm, which still needs to be addressed.

5 Conclusions and future work

This work proposed an integrated hybrid framework for aspect-based opinion mining, with an emphasis on aspect-sentiment detection and extraction, aspect-sentiment classification, and aspect-based summary generation. We obtained classification results with improved precision (0.85) when compared to the alternative methods available. The proposed method is quite generalised and it can classify aspect-based opinions in multiple domains.

The proposed method does feature a number of limitations that must be borne in mind when interpreting the findings. First, the proposed method lacks the capacity for implicit aspect extraction. Second, it exhibits low accuracy in terms of sentiment classification due to the lack of consideration of informal opinion carriers such as emoticons and slang. Third, the measure adopted for the grouping of similar aspects is based on the semantic similarity approach, whereas adopting the semantic relatedness measure could bring about promising results with respect to aspect grouping. Finally, the proposed method does not address multiple aspects and the related sentiments that are present in a single sentence. For example, in the sentence “The touchscreen is smooth but the battery life is not good”, “touchscreen” and “battery life” are the two aspects, while “smooth” and “not good” are the opinions expressed regarding those two aspects, respectively.

The possible future directions of this work can be summarised as follows. First, the pattern-based approach generates several redundant candidate terms for the aspects and their associated opinions. To eliminate such redundant terms, an efficient post-processing algorithm is required. Second, it would be valuable to determine a means of extracting implicit aspects from user reviews. Third, the proposed method could be extended to include informal opinion carriers such as emoticons and slang, which would enhance the accuracy of its sentiment classification. Fourth, the value of related measures such as WordNet and latent semantic analysis could be investigated for the grouping of similar aspects. Finally, the ability to handle sentences that contain multiple aspects and their related sentiments could improve the efficiency of the proposed system.

References

1. Asghar, M.Z., Ahmad, S., Qasim, M., Zahra, S.R., Kundi, F.M.: SentiHealth: creating health-related sentiment lexicon using hybrid approach. SpringerPlus. **5**(1):1139 (2016)
2. Asghar, M.Z., Khan, A., Ahmad, S., Kundi, F.M.: Preprocessing in Natural Language Processing in Emerging Issues in the Natural and Applied Sciences, Baku, Azerbaijan: Progress IPS LLC, pp. 52–161 (2013)
3. Asghar, M.Z., Khan, A., Ahmad, S., Khan, I.A., Kundi, F.M.: A unified framework for creating domain dependent polarity lexicons from user generated reviews. PLoS ONE **10**(10), e0140204 (2015)
4. Asghar, M.Z., Khan, A., Ahmad, S., Qasim, M., Khan, I.A.: Lexicon-enhanced sentiment analysis framework using rule-based classification scheme. PLoS ONE **12**(2), e0171649 (2017). doi:10.1371/journal.pone.0171649
5. Bird, S., Klein, E., Loper, E.: Natural Language Processing with Python. O'Reilly Media, Inc (2009)
6. Chinsha, T.C., Shibily, J.: A syntactic approach for aspect based opinion mining. In: Proceedings of the IEEE 9th International Conference on Semantic Computing (2015)
7. Esuli, A., Sentiwordnet, S.F.: A publically available lexical resource f or opinion mining. In: Proceedings of LREC, vol. 6 (2006)
8. Htay, S.S., Lynn, K.T.: Extracting product features and opinion words using pattern knowledge in customer reviews. Sci. World J. **3**, 2013 (2013)
9. Hu, M., Liu, B.: Mining Opinion Features in Customer Reviews. American Association for Artificial Intelligence (2004)
10. Kennedy, Alistair, Inkpen, D.: Sentiment classification of movie reviews using contextual valence shifters. Comput. Intell. **22**(2), 110–125 (2006)
11. Kessler, L.S., Nicolov, N.: Targeting sentiment expressions through supervised ranking of linguistic configurations. In: ICWSM (2009)
12. Khan, K., Baharudin, B., Khan, A.: Identifying product features from customer reviews using hybrid patterns. Int. Arab. J. Inf. Technol. **11**(3), 281–286 (2014)
13. Kreutzer, J., Witte, N.: Opinion Mining Using SentiWordNet, pp. 1–19 (2013)
14. List of intensifiers https://archive.org/stream/intensifiersincu00benz/intensifiersincu00benz_djvu.txt
15. Liu, B.: Sentiment analysis and opinion mining. Synth. Lect. Human Lang. Technol. **5**, 1–167 (2012)
16. Maharani, W., Widyantoro, D.H., Khodra, M.L.: Aspect extraction in customer reviews using syntactic pattern. Procedia Comput. Sci. **59**(Iccsci), 244–253 (2015)
17. Maharani, W., Widyantoro, D.H., Khodra, M.L.: Learning-based aspect identification in customer review products. In: The 5th International Conference on Electrical Engineering and Informatics 2015 August 10–11, Bali, Indonesia, pp. 71–76 (2015)
18. Maharani, W., Widyantoro, D.H., Khodra, M.L.: SAE: syntactic-based aspect and opinion extraction from product reviews. ICAICTA 2015-2015 Int. Conf. Adv. Informatics Concepts, Theory Appl. (2015)
19. Manek, A.S., Shenoy, P.D., Mohan, M.C., Venugopal, K.R.: Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. World Wide Web **20**, 135 (2017). doi:10.1007/s11280-015-0381-x
20. Marrese-Taylor, E., Velásquez, J.D., Bravo-Marquez, F.: A novel deterministic approach for aspect-based opinion mining in tourism products reviews. Expert Syst. Appl. **41**(17), 7764–7775 (2014)
21. Pang, B., Lee, L.: A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics (AMACL'04), pp. 271–277 (2004)
22. Parkhe, V., Biswas, B.: Sentiment analysis of movie reviews: finding most important movie aspects using driving factors. Soft Comput. **20**(9), 3373–3379 (2016)
23. Popescu, A.-M., Etzioni, O.: Extracting product features and opinions from reviews. Natural Language Processing and Text Mining, pp. 9–28. Springer, London (2007)
24. Provost, F.J., Fawcett, T., Kohavi, R.: The case against accuracy estimation for comparing induction algorithms. In: Proceedings of the Fifteenth International Conference on Machine Learning **98**, 445–453 (1998)
25. Qiu, G., Liu, B., Bu, L., Chen, C.: Opinion word expansion and target extraction through double propagation. Comput. Linguist. **37**(1), 9–27 (2011)
26. Rana, T.A., Rana, T.A.: Exploiting Sequential Patterns to Detect Objective Aspects from Online Reviews Exploiting Sequential Patterns to Detect Objective Aspects from Online Reviews, no. August (2016)
27. Samha, A.K., Li, Y., Zhang, J.: Aspect-based opinion extraction from customer reviews, arXiv preprint arXiv:1404.1982 (2014)
28. Singh, V., Piryani, R., Uddin, A., Waila, P.: Sentiment analysis of movie reviews: a new feature-based heuristic for aspect-based sentiment classification. In: Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), International Multi-Conference, IEEE, pp. 712–717 (2013)
29. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. Comput. Linguist. **37**(2), 267–307 (2011)
30. Thet, T.T., Khoo, C.S., Shakthikumar, S.: Sentiment analysis of movie reviews on discussion boards using a linguistic approach. In: Proceedings of the 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion, pp. 81–84. ACM (2009)
31. Turney, P.D.: Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational linguistics, pp. 417–424, July (2002)
32. Varghese, R., Jayasree, M.: Aspect based sentiment analysis using support vector machine classifier. In: Advances in Computing, Communications and Informatics (ICACCI), International Conference on IEEE, pp. 1581–1586 (2013)
33. Zhai, Z., Liu, B., Xu, H., Jia, P.: Clustering product features for opinion mining. In: Proc. fourth ACM Int. Conf. Web search data Min.-WSDM '11, no. March, p. 347 (2011)
34. Zhuang, L., Jing, F., Zhu, X.: Movie review mining and summarization. In: Proceedings of the 15th ACM International Conference on Information and Knowledge Management, pp. 43–50 (2006)

Muhammad Zubair Asghar is an HEC approved Ph.D. supervisor recognized by Higher Education Commission (HEC), Pakistan. His Ph.D. research includes recent issues in Opinion Mining and Sentiment Analysis, Computational Linguistics and Natural Language Processing. He is working as Assistant Professor in the Institute of Computing and Information Technology, Gomal University, D.I. Khan, Pakistan. He has authored more than 40 publications in journals of international repute (JCR and ISI indexed) and having more than 20 years of University teaching and laboratory experience in Expert Systems, Neural Networks, Python, Visual Prolog and Assembly language.

Aurangzeb Khan did Master in Computer Science from The University of Peshawar, Pakistan, and Ph.D. in Information Technology from Universiti Teknology, PETRONAS, Malaysia. Currently, he is Associate Professor in Computer Science/Chairperson, Department of Computer Science, University of Science and Technology, Bannu Pakistan; He also holds the charge of Director Academics. As a researcher, he has more than 50 publications, and is currently working on two Sponsored Research Projects in Information Technology.

Syeda Rabail Zahra is a research scholar at Gomal University. Currently she is working as a Lecturer at ICIT, Gomal University. Her research focus is on aspect-level sentiment analysis and its applications. Other topics of interest include data mining techniques for opinion mining in customer reviews and Artificial Intelligence.

Shakeel Ahmad received his B.Sc. with distinction from Gomal University, Pakistan (1986) and M.Sc. (Computer Science) from Qauid-e-Azam University, Pakistan (1990). He received his Ph.D. degree in computer science in January 2008 and completed one year post-doctoral study from University Science Malaysia (USM) in 2010. He started his career as a lecturer in 1990 and served for 11 years in Institute of Computing and Information Technology (ICIT), Gomal University Pakistan. Then he served as Assistant Professor, Associate Professor, Professor and Director Institute of Computing and Information Technology (ICIT), Gomal University Pakistan during 2001–2014. Now days, he is serving as professor in Faculty of Computing and Information Technology at Rabigh (FCITR), King Abdulaziz University Jeddah, Kingdom of Saudi Arabia. Dr. Shakeel has an outstanding teaching career with proficient research background, reflecting more than 27 years of teaching and research experience in Performance modelling, Sentiment Analysis and Text Mining, Optimization of congestion control techniques and Electronic learning. He has produced many publications in Journal of international repute and presented papers in International conferences.

Fazal Masud Kundi received his MS-IT from University of Peshawar and Ph.D. in Computer Science from Gomal University. He is working as Assistant Professor in the Institute of Computing and Information Technology, Gomal University, D.I. Khan, Pakistan. His research interests include recent trends in data mining, text mining and computational linguistics. He has more than 40 publications in journals of international repute. He has vast laboratory experience in different research tools and languages, such as NLTK-based python, Weka, Rapid Miner, Excel Miner, R, and now, he is working on KNIME analytics platform.