SN Computer Science (2020) 1:128
https://doi.org/10.1007/s42979-020-00138-7

ORIGINAL RESEARCH

A Hybrid Co-occurrence and Ranking-based Approach for Detection of Implicit Aspects in Aspect-Based Sentiment Analysis

M. Devi Sri Nandhini[1] · G. Pradeep[2]

Received: 27 March 2020 / Accepted: 31 March 2020

Abstract

People are increasingly using Web sites and Web services to express their opinions since the inception of Internet. Sentiment analysis is an active research domain that aims at the extraction of sentiments or opinion from the user text, thereby getting the associated sentiment orientation. Although sentiments can be analyzed at document level and phrase level, unfortunately these options are not sufficient for fine-grained analysis of sentiments. Therefore, it makes sense to focus on aspect-based sentiment analysis which is very promising in terms of providing accurate predictions on user sentiments. There is a lot of scope for the research community to provide solutions for various challenges involved in performing sentiment analysis at the aspect level. The primary goal of this work is to extract the implicit aspects from opinionated document using the co-occurrence of aspects with feature indicators and ranking the pair based on their frequency of co-occurrence. As a first step toward achieving this objective, a novel algorithm is proposed to detect the implicit aspects through co-occurrence and ranking. The proposed algorithm is reliable as it uses the association between explicit aspects and sentiment words to detect implicit aspects.

Keywords Implicit aspects · Feature indicators · Co-occurrence

Introduction

There have been a lot of technological advancements happening across the Web, and people are increasingly engaging themselves in online activities such as discussion forums, chatting and social media, providing comments in product Web sites that are capable of serving as vital sources for many research initiatives. Sentiment analysis domain is getting benefited largely by such technological improvements. It can be defined as the systematic way of detecting human emotions toward various entities. As said earlier, the source for carrying out this analysis is the Internet activities carried out by people such as posting their opinions for services, politics and movies. From the application perspective, sentiment analysis can be used in computational advertising and market intelligence that finds the extent of user satisfaction regarding products and services which would provoke the initiatives toward alleviating the weaknesses of products and services. Sentiment analysis also finds its application like analyzing political scenario and predicting the election results. With respect to the online reviews and social media content, the sentiment analysis technique proves to be useful in finding out the spikes in sentiment. Whenever negative threads emerge in social media about a particular product or service, it is possible to detect those threads and take proactive measures to deal with it. Sentiment analysis has several applications in day-to-day activities; sentiment analysis is mainly used to examine the opinions of general public regarding the issues in political sector. It finds wide application in the field of marketing intelligence; in this area, it can aid in finding the level of satisfaction attained by users on products or services, thereby offering its help

This article is part of the topical collection "Advances in Computational Approaches for Artificial Intelligence, Image Processing, IoT and Cloud Applications" guest edited by Bhanu Prakash K N and M. Shivakumar.

✉ M. Devi Sri Nandhini
nandhini.avcce@gmail.com

G. Pradeep
pradeep.g8@gmail.com

1 A.V.C College of Engineering, Mannampandal, Mayiladuthurai, Tamilnadu, India

2 Department of Computer applications, A.V.C College of Engineering, Mannampandal, Mayiladuthurai, Tamilnadu, India

Published online: 22 April 2020

in overcoming the weaknesses. Sentiment analysis can be utilized for estimating the price changes according to recent trends. It can serve as a valuable aid in designing new products and services as per customer expectations. Through accurate analysis of customer reviews, sentiment analysis can be helpful in enhancing the product features. In general, a larger population of people decide about buying products and using services based on these reviews.

Sentiment analysis can be categorized into three types of analysis, namely document-based analysis, sentence-based analysis and aspect-based analysis. With respect to document-based analysis, a complete review of the user is taken as a single unit and it facilitates to predict only the overall sentiment associated with the review. The main flaw in the document-level analysis is that it will not take into account about each opinionated sentence. To overcome this drawback, sentence-level analysis can be carried out which categorizes the sentences to be objective and subjective. Only subjective sentences contain opinions. Therefore, it analyzes the subjective sentences and brings out the opinions expressed at the sentence level. So, in comparison with document-level analysis, the sentence-level analysis is better. A far more accurate and fine-grained analysis is possible by analyzing at the aspect level. With respect to aspect-based analysis, opinion associated with each and every aspect in the review can be determined. So, aspect-based analysis gives very accurate predictions. Figure 1 presents the categories in sentiment analysis.

With respect to the document-based analysis, it is necessary to find out the opinion words present in the full review text. The review text may contain a detailed content or a brief content. This level of analysis predicts the sentiment of the overall document omitting the task of finding the sentiment of each aspect. In this type of sentiment analysis, the result pertains to the overall sentiment of the review text. In the document-based analysis, the entire review is taken as one unit and based on certain opinion words, it is concluded that the overall opinion associated with the review indicates positivity or negativity or neutrality. This kind of analysis is very critical in the areas of social research and psychological learning in social networks. It is also useful in analyzing the records of patients who are undergoing medical treatment.

From the sentence-level standpoint, the aim lies in determining the sentiment associated with the sentence. This approach will not examine each aspect as a discrete instance. In this approach, the primary task is to examine the nature of each sentence. If it is a subjective phrase, then it is a potential candidate which has the grounds to possess some user sentiment in it. On the other hand, if it is an objective phrase, then it need not be examined at all. Such subjective sentences are treated as small documents, and the opinion associated with those sentences are found. Sentence-level sentiment analysis is usually affected by the atmosphere of the sentence. It is critical for examining the twitter messages, Facebook updates and brief messages.

Eventually, the aspect-based analysis which otherwise called as feature-based analysis is a fine-grained approach in sentiment analysis. It involves the task of finding the sentiment expressed by a user regarding a certain aspect of a product or an entity. To carry out a sentiment analysis task based on aspects, it is required to bring out the entities and their associated aspects. Then, the sentiment associated with a given aspect is determined. As the last step of aspect-based sentiment analysis, it is possible to obtain the summarized

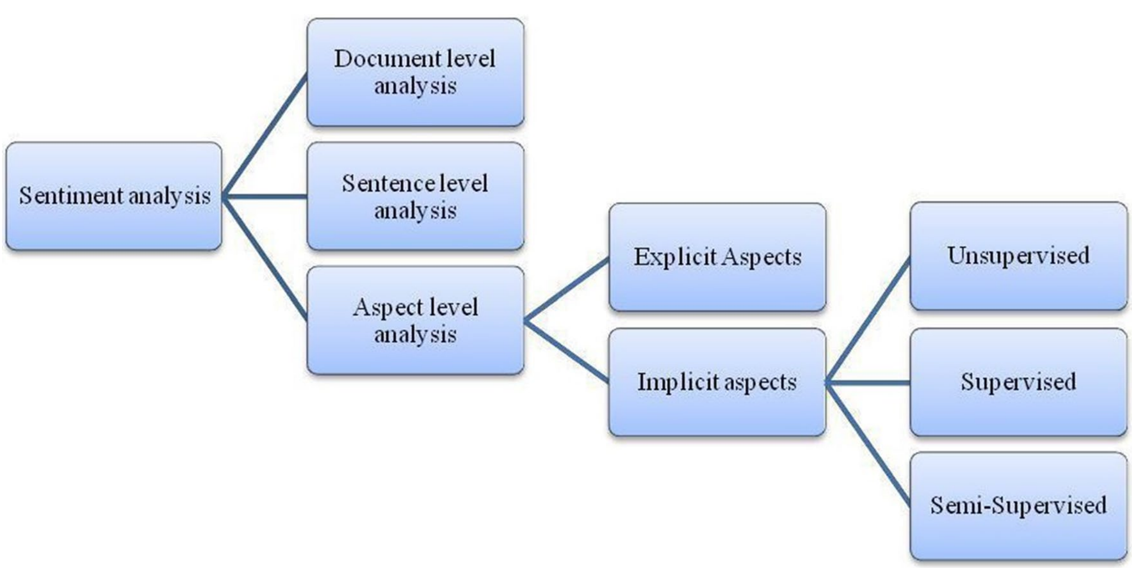


Fig. 1 Categories of sentiment analysis

opinion and its corresponding visualization results. The aspect that is extracted can be either explicit or implicit. The extracted aspect is taken to be explicit if it is openly stated in the review sentence. If not, it is decided to be an implicit aspect.

Role of Implicit Aspects in Aspect-Based Sentiment Analysis

With respect to aspect-based sentiment analysis, it becomes crucial to determine the polarity associated with different features or aspects of the entities under evaluation. There are two types of aspects that are prevalent in user-generated content, namely explicit and implicit. Aspects which are directly stated in the subjective phrases are regarded as explicit. It is relatively easier to extract when compared to implicit aspects that are hidden in the opinionated text. The implicit aspects have to be identified from the clues available in the sentences, and this is the main issue that gains focus with respect to the phrases that involves implicit aspects. For example, if a user comments as: The gadget is affordable, then he/she actually refers to the price aspect, even though it is not directly specified. In another example, the user mentions "the phone is glossy, appealing and it is easy to operate and so much comfortable to scroll the menus." In this case, the terms glossy and appealing are the clues that denoted the implicit aspect appearance, whereas the phrase "easy to operate" refers to the implicit aspect of functionality.

Figure 2 presents the kinds of clues that denote implicit aspects.

Thus, the general strategy to find implicit aspects involves first identifying the clues in the opinionated document and mapping those clues to the appropriate implicit aspects. From the above example, it is concluded that the clues can be either a single word such as "glossy" or it can be an expression involving a group of words, for example "easy to operate." Sometimes, the same implicit word might be referred to by different words in different phrases. In one sentence, user might have mentioned the gadget is cheap, later somewhere in the review he/she might have again mentioned that the gadget is affordable. Both the terms, affordable and cheap, refer to the implicit aspect price. Such issues need to be handled in this stage of analysis.

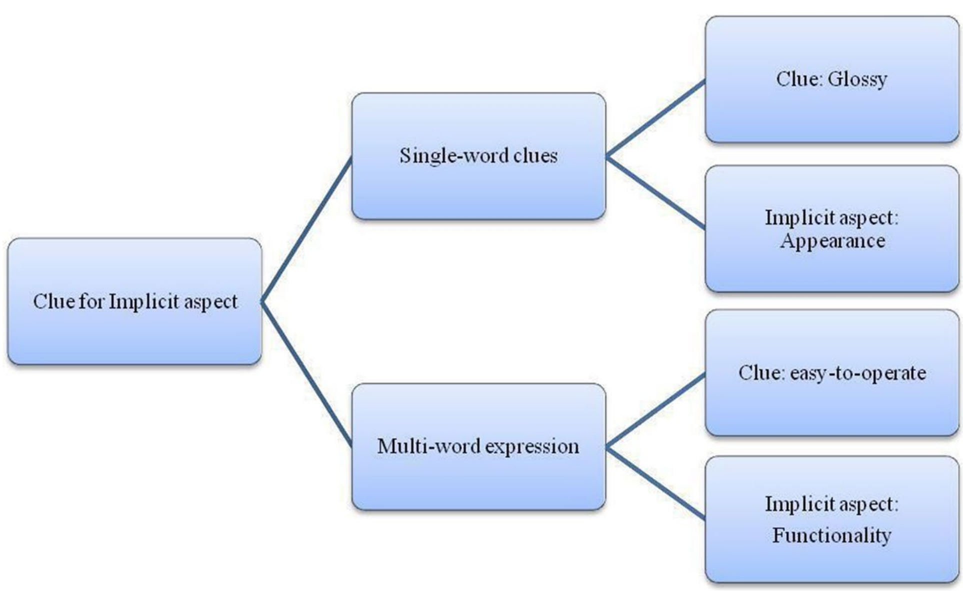
Also it is critical to learn the context dependent meaning of the opinion word. As an example, consider two sentences: "The screen resolution is high," and "The phone is of high price." The opinion word "high" expresses a positive opinion for screen aspect in the former sentence, and the same word expresses a negative opinion for the price aspect in the later sentence. Such issues can be resolved by carrying out context-dependent analysis of opinion words.

Related Work

In their work, the researchers, Vo et al. [1], have proposed a new approach to find and summarize various entity features and their associated views. This summary of viewpoints is generated by analyzing several reviews about products in a specific area. Tools that are available for working with natural language are used to bring out the facts related to the syntaxes. Those tools also aid in pulling out the relations between views and features that are hinted indirectly. It is possible to automatically obtain the coarse knowledge from reviews through co-references.

A new extraction method based on frame of reference was proposed by Sun et al. [2]. In this method, implicit features

Fig. 2 Types of clues denoting implicit aspects



are extracted depending upon the likeness. From the context of hidden features, the similarity between the features of the entity is taken into account. This method involves three steps. The first step is to build a rectangular array of quantities to indicate the interconnection that links the words expressing views and aspects. The second task is to look for an implicit aspect, if found, generate a list of possible implicit aspects. The third task is to determine the exact implicit aspects using the computed points of the implicit aspects. These points are computed by utilizing the words expressing user views and the surrounding information of the implicit aspects. In this case, context information and opinion words are used to extract the implicit aspects. This work suffers from few pitfalls, for example, in the mini compilation of texts, it fails to determine the hidden aspects accurately.

The idea of applying Naïve Bayes method to identify aspects was proposed by the Mohamad syahrul et al. [3]. The authors have suggested the idea of using Naïve Bayes to detect aspects from the reviews. This approach also aids in categorizing the aspect's polarity. In this approach, the task is carried out in three phases, namely preprocessing, feature classification. Data preprocessing is done by performing the subtasks such as removing the commonly used words that are excluded from searches, breaking up the text into discrete units and reducing a word to its root. Parts-of-speech tagging method and the Goodness-of-fit method can be utilized for choosing the features. Two-variable classification is done by using Naïve Bayes model. The usage of Chi Square method has proven in stepping down the computation time required for classifying the variables. But it has brought down the performance of the proposed approach. During the classification process, documents that could not be categorized by the proposed approach resulted in improper classification. When an important feature is missing from previous processes, misclassification is caused.

Statistical methods were utilized by Suresh and Muthukumaran [4] in extracting implicit aspects. The authors have combined statistical methods with traditional linguistic rules and representations. They have proposed a sentiment analyzer which extracts the sentiments associated with a subject from online text documents. In order to analyze the polarity in online product reviews, Naive Bayes classification model and Hidden Markov Models were applied. These two models were chosen because of the inherent computational simplicity and stochastic robustness. First, the process starts by preprocessing the review document from the datasets. The next task is to extract the sentences according to the delimiters. Further, the process proceeds by removing stop words and words with frequency less than three. The next step is to form a feature vector to represent review documents. LDMA and the original LDA models are applied to discover more informative words for aspects or topics.

Tubishat et al. [5] proposed a hybrid approach in their work to extract implicit aspects using corpus, dictionary, co-occurrence and Web-based similarity.

Fei et al. [6] detected implicit aspects in their work using dictionary. This dictionary was compiled by choosing several opinion words along with its glosses from online dictionaries.

Rana et al. [7] extracted implicit aspects using co-occurrence approach and normalized Google distance to see the extent to which a given candidate implicit aspect matches with a opinion word.

Brun et.al. [8] proposed an integrated approach using a parser, lexicons and SVM approach to carry out the detection of aspects.

Hai et.al. [9] proposed an approach involving two phases using co-occurrence and association rule mining to detect implicit aspects.

Kiritchenko et.al. [10] detected aspects and sentiments using supervised approach aided by word-aspect association lexicons.

Popescu et.al. [11] proposed an unsupervised extraction system to detect fine-grained aspects.

Taxonomy of Implicit Aspect Extraction

For performing hidden aspect extraction, there are three kinds of approaches depending upon the method of extraction, namely unsupervised approaches, supervised approaches and semi-supervised approaches. Figure 3 presents a brief summary of each of these approaches.

Approaches for Implicit Aspect Extraction Unsupervised Approaches

In order to detect implicit aspects from the review documents, unlabeled data are used by unsupervised approaches. Unsupervised approaches do not involve any training process. According to the literature review carried out, commonly used unsupervised methods are dependency parsing, hierarchy, co-occurrence, clustering and rule-based Ontology. Few unsupervised approaches and the ideas utilized in those approaches are presented below:

Syntactic or Dependency Parsing

It is widely utilized in sentiment analysis. Dependency parsing is the task of extracting a dependency parse of a sentence that represents its grammatical structure and defines the relationships between "head" words and words, which modify those heads. Syntactic parsing or dependency parsing is the task of recognizing a sentence and assigning a syntactic structure to it.

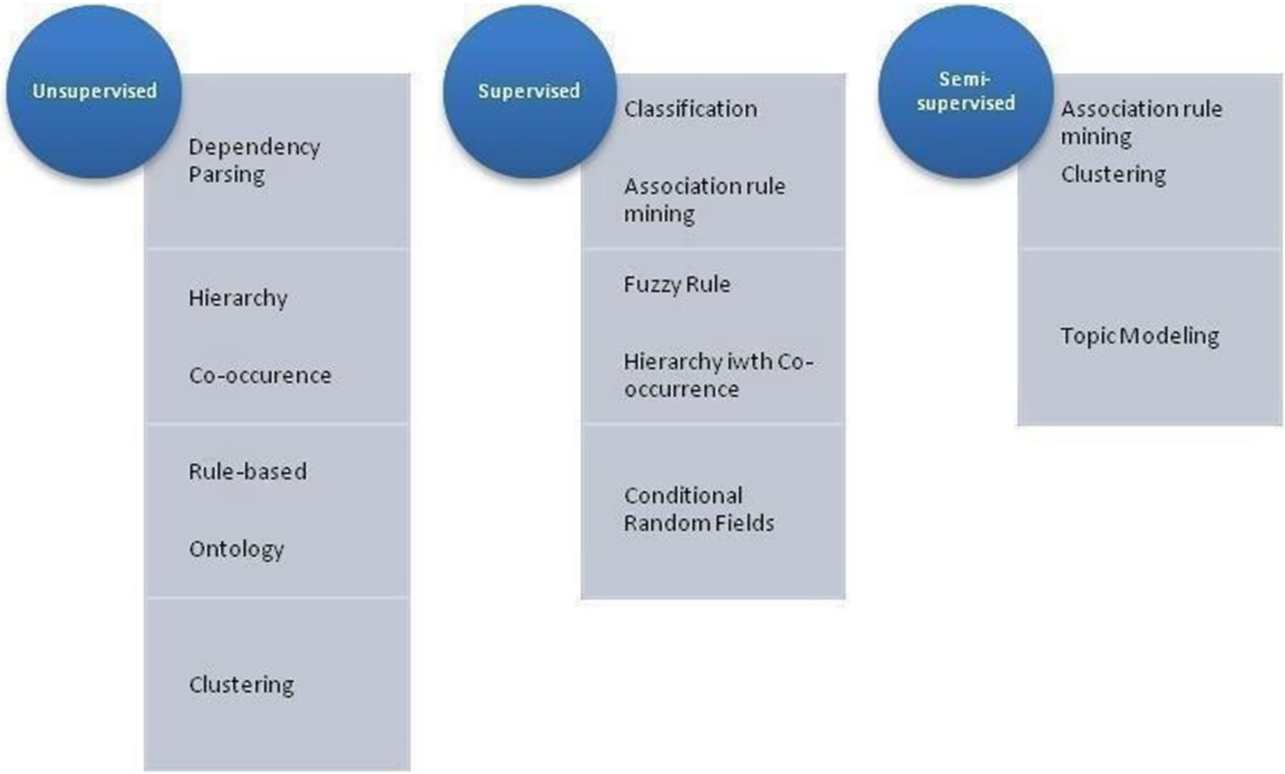


Fig. 3 Implicit aspect extraction approaches

Hierarchy

In some of the existing works, hierarchy was used for extracting implicit aspects. To create the used hierarchy, first, it is necessary to fetch the primary entity hierarchy from the Web site corresponding to the specific product. In the hierarchy, for every node associated with a product feature, attach those review texts in which the exact features are stated and for which views were expressed. Next, convert these linked review text into a vector. This vector contains the complete list of sentiment words in the review text.

The next task is to calculate the centroid for each aspect. This is done by finding the mean value of the sentiment words present in the vector. The next step is to calculate the cosine of the angle between two vectors projected in a multi-dimensional space. It is a measure of similarity between two nonzero vectors. At last, the aspect that has the highest similarity value associated with it will be picked up as the right hidden aspect.

Co-occurrence

For extracting hidden aspects, one of the promising ideas that would work out well is to make use of the degree of coexistence of the words expressing user views and the corresponding entity on which the views are expressed. An enhanced rectangular array of coexistence or co-occurrence can be used for hidden aspect extraction. Double propagation technique can be used to extract the explicitly stated aspects and their associated view-expressing words. Furthermore, the co-occurrence matrix is built by using the extracted feature and opinion terms. This matrix incorporates the recurrence of coexistence between the explicitly stated features and their opinion terms. In addition, it contains the recurrence of coexistence between the extracted explicitly stated aspects and the hypothetical or speculative words. Finally, the co-occurrence matrix is searched for every view-expressing word that does not have an explicitly stated aspect; it is looked up for finding the aspects that coexisted with the given opinion term. Depending on the potential hidden aspects which were fetched so far, that feature which prominently coexisted with other speculative words in the phrase is selected.

In an alternative approach, it is also possible to make use of a graph for the extraction process. The graph is constructed by using the coexistence relationship between the explicitly stated aspects and the associated opinion terms. Every edge of the graph is given a label by specifying the total number of coexistences of their aspect and the opinion term. A weight is affixed to the coexistence value. The value of weight denotes the extent of attachment between the specific aspect and the opinion term. Mapping is done with the graph, for every opinion term that does not have a corresponding explicit feature. Finally, that aspect in the pair which has the exact opinion term with the topmost edge value is taken as a hidden aspect.

Supervised Approaches

In order to detect implicit aspects from reviews, the supervised approaches make use of labeled data; these approaches make use of any algorithm that requires training. According to the survey conducted for this work, commonly used supervised methods for implicit aspect detection includes association rule mining, classification, conditional random fields, hierarchy with co-occurrence and fuzzy rule. Some of the supervised approaches and the ideas utilized in those approaches are presented below:

Classification

The classification approach encompasses the usage of Naïve Bayes classifier, WordNet and corpus as a hybrid technique for extracting implicit aspects. The first step is to extract all the adjectives from the corpus. Next, WordNet is used to extract the words which have lexical relationships with the specified adjectives. For a given word, its synonyms and antonyms are said to have a lexical relationship with that particular word. The final step is to use the pulled out data to upskill the specific classifier for extracting the hidden aspects.

In yet another classification technique, it is recommended to extract the explicitly specified features and their associated opinion terms. Then, for every opinion term, group all features that coexist with the given opinion term into a single group. The subsequent task is to create a training document for every feature-view couple in the specified group. This training report contains the phrases that incorporate the identical couple in the dataset. Then, an appropriate classification technique is used for mapping the opinion term that does not have an explicit aspect.

Dictionary-based approach can be used to find out those aspects that are implicitly denoted by the opinion words that occur as adjectives in the sentences. Then, find all glosses from five online English dictionaries for every opinion word. The next task is to find out the nouns which are treated as probable hidden aspect nominees. At last, using these lexical links, the adjective-related nouns are classified and detected.

Association Rule Mining

In this approach, the idea is to find out from the dataset, those opinion terms that do not have an associated explicit aspect and label those words as implicit features. Association rule mining is a technique to detect patterns in data. It detects the features which co-occur and the features which are correlated. The primary task is to form the association rules. Some of the measures of effectiveness of the rule includes support, confidence and lift.

Hierarchy with Co-occurrence

This approach insists to perform manual labeling. Those opinion terms that does not have an associated explicit aspect are manually labeled with their hidden aspects. In this approach, the main task is to extract the phrases with stamped hidden aspects and group them as per the opinion words. The next task is to enumerate the possible hidden aspect candidates through appropriate searching procedure. Further, calculate a score for each hidden aspect candidate. This score is calculated by using the extent of coexistence between the opinion term and other terms in the equivalent sentence. As a subsequent step, choose the implicit aspect candidates pertaining to a higher score in comparison with a given threshold. Then, the double propagation technique is applied to pull out the explicit aspects in their hidden contexts. Moreover, check if the candidate hidden aspect has a link with the pulled out explicit aspect, if so, then it is taken to be a correct implicit aspect.

Semi-supervised Approaches

These approaches use labeled as well as unlabeled data to detect implicit aspects; otherwise, these approaches involve little training. As per the literature survey conducted for this work, widely applied semi-supervised approaches for extracting hidden aspects are clustering and topic modeling.

Problem Formulation and Proposed Work

Let D denote a document that contains the sentences present in a user review:

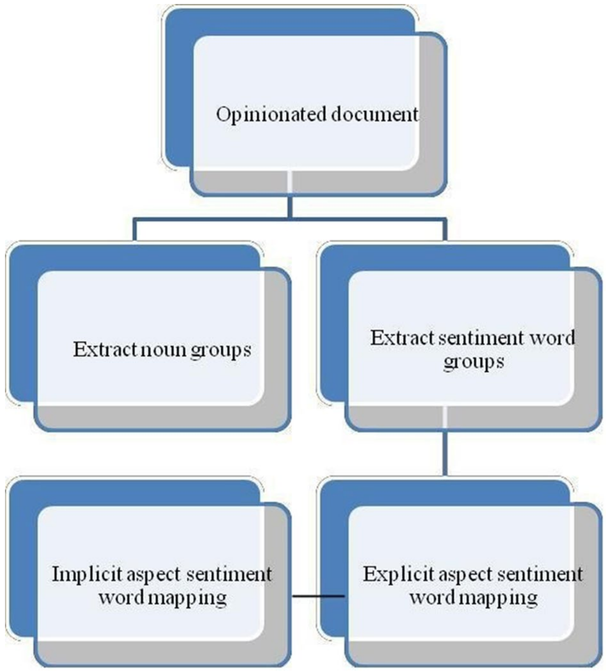


Fig. 4 Implicit aspect extraction as a consequence of explicit aspect-polarity mapping

$$D = \{s_1, s_2, s, \dots, s_n\} \tag{5.1}$$

where s_1, s_2,…,s_n denote the sentences contained in the document D.

Each sentence in D, s_i can be either an opinionated sentence (os) or a non-opinionated sentence (nos). Let OS denote the set of all opinionated sentences. Some of the sentences in OS may not contain explicit aspects, which poses a challenge for finding out the hidden aspects. Let NOS denote the set of all non-opinionated sentences which need not be subjected to analysis.

Let E be an entity that has an aspect "a" in an opinionated text. The sentences in the document may express positivity, negativity or intermediate sentiment on aspect "a."

Entity E is thus a set of aspects, A as given below:

$$A = \{a_1, a_2, \dots, a_n\} \tag{5.2}$$

This set of aspects may include the entity E itself as a special aspect. Each aspect aid in A is denoted as a group of words or phrases:

$$W_i = \{W_{i,1}, W_{i,2}, \dots, W_{i,n}\} \tag{5.3}$$

Each element in the above set is a set of synonyms for each corresponding aspect. The proposed work is to extract implicit aspects from the opinionated document using co-occurrence and ranking approach. The fundamental concept is that generally adverbs and adjectives indicate features. Such feature indicators serve as the clues using which implicit aspects can be detected. Some of the adverbs and adjectives are commonly used as feature indicators such as nice, good, worst or bad. Such indicators can be directly mapped with their respective aspects.

Let k be the set of feature indicators that can be used to detect implicit aspects. It is denoted by:

$$k = \{k_{i,1}, k_{i,2}, \dots, k_{i,n}\} \tag{5.4}$$

In the above set of feature indicators, each element represents a set of adverbs and adjectives that acts as clues for detecting the hidden aspects.

Implicit aspects can be found by making use of feature indicators or sentiment words. As shown in Fig. 4, after performing explicit aspect-sentiment word mapping, the extraction of implicit aspects is carried out. Based upon the past observations, it is possible to determine that an aspect is implied by a feature sentiment word or not. It can be based on a previously collected text corpus pertaining to the domain based on how frequently a feature indicator co-occurs with a specific aspect. If a feature indicator appears in a opinionated sentence, then it is taken that it implies that specific aspect with which it co-occurred for the highest number of times.

A novel algorithm for extracting implicit aspects through co-occurrence and ranking

Procedure Extracting implicit aspects

Require sentence is a data structure that contains set of all opinionated sentences in the document

```
1: nouns_list={}
2: explicit_aspect_list={}
```

#Finding explicit aspects from opinionated sentences

```
3: for each noun n in sentence do
4:     add n to nouns_list
5:     add n to explicit_aspects_list
```

#Building Feature indicators or Sentiment words list

```
6: for each adverb x in sentence do
7:     add x to sentiment_word_list
8: for each adjective x in sentence
9:     add x to sentiment_word_list
```

#Creating a mapping between each explicit aspect and its co-occuring sentiment word

```
10: for each explicit aspect e in explicit_aspect_list do
11:     for each sentiment word sw in sentiment_word_listdo
12:         if co-occurrence(e,sw) is true then:
13:             Add the explicit aspect(e) and sentiment-word(sw) pair to the map, M
13:             Update the count for co-occurrence of e and sw by unit increment
14: Rank the sentiment word-explicit aspect pairs based on their frequency
of co-occurrence.
```

 # Implicit aspect detection

```
15: for each opinionated sentence (os) that has no explicit aspectdo:
16:         for each sentiment word sw in os do:
17:                 Find from map, the explicit aspect, say i with which the
                    sentiment word has co-occurred for the highest number of times
18:                 Finalize i, as the implicit aspect implied by the sentimentword.
19:                 Append implicit aspect, i, to the implicit_aspect_list
20: return implicit_aspect_list
```

Conclusion and Future Work

In this paper, a basic algorithm is proposed for extracting implicit aspects through co-occurrence and ranking techniques. It provides the sequence of actions to be carried out on the way to find hidden aspects. The importance of aspect-based sentiment analysis together with the contribution of implicit feature detection in producing fine-grained analysis of user sentiments is discussed in-depth. Various methods of extracting implicit aspects such as supervised, unsupervised and semi-supervised approaches have been explored. It is a promising domain that has a lot of challenges to be tackled by the research community. A reasonable improvement can be expected in the overall performance of sentiment analysis if the subproblems are addressed appropriately. Future research involves dealing with irony sentences and intensifiers that may strengthen or weaken the sentiment orientation of a given aspect that demand in-depth analysis.

Acknowledgements We would like to extend our sincere thanks to the authors of the reference papers for their valuable ideas and the recommended methods in the area of sentiment analysis. We also thank the reviewers for their useful comments and suggestions.

References

1. Vo AD, Nguyen QP, Ock CY (2018) Automatic knowledge extraction for aspect-based sentiment analysis of customer reviews. In: ICCMS, 2018. Sydney, Australia.
2. Sun L, Li S, Li J, Lv J (2014) A novel context-based implicit feature extracting method. In: 2014 International conference on data science and advanced analytics. https://doi.org/10.1109/dsaa.2014.7058106.
3. Mubarok MS, Adiwijaya, Aldhi MD. Aspect-based sentiment analysis to review products using Naïve Bayes. AIP Conf Proc. 2017;1867:020060.
4. Suresh P, MuthuKumaran S. Sentiment analysis of product reviews using LDA method based on customer text content. Int J Comput Sci Eng Technol. 2015;5(12).
5. Tubishat M, Idris N. Explicit and Implicit Aspect Extraction using Whale optimization algorithm and Hybrid approach. In: Atlantis Highlights in Engineering, IcoIESE 2018, Volume 2. 2018.
6. Fei G, Liu B, Hsu M, Castellanos M, Ghosh R. A dictionary-based approach to identifying aspects implied by adjectives for opinion mining. In: 24th international conference on computational linguistics. p. 309, 2012.
7. Rana TA, Cheah YN. Hybrid rule-based approach for aspect extraction and categorization from customer reviews. In: IT in Asia (CITA), 2015 9th international conference on IEEE, p. 1–5, 2015.
8. Brun C, Popa DN, Roux C. XRCE: Hybrid classification for aspect-based sentiment analysis. In: Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014) (Association for Computational Linguistics and Dublin City University, Dublin, Ireland, 2014), p. 838–42.
9. Hai K, Chang, Kim J. Implicit feature identification via co-occurrence association rule mining. In: Computational linguistics and intelligent text processing. p. 393–404.
10. Kiritchenko S, Zhu X, Cherry C, Mohammad S. NRC-Canada-2014: detecting aspects and sentiment in customer reviews. In: Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014) (Association for Computational Linguistics and Dublin City University, Dublin, Ireland, 2014), p. 437–42.
11. Popescu A-M, Etzioni O. Extracting product features and opinions from review HLT/EMNLP 2005, 2005.