INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING 2019, ICRTAC 2019

Implicit Aspect Extraction for Sentiment Analysis: A Survey of Recent Approaches

Vaishali Ganganwar, R.Rajalakshmi*

Vellore Institute of Technology, Chennai, Tamil Nadu, India

Abstract

The research in Sentiment analysis (SA) is in vastly growing stage as people become more expressive on social media, blogs, forums and e-commerce websites by sharing their opinions, reviews and comments. In Aspect-level SA opinions about various aspect or features of an entity is extracted. Users specify aspects by explicit words (i.e. Explicit aspects) or sometimes the aspects must be inferred from the text (implicit aspects).Detecting implicit aspects is challenging but very important and limited studies focused on the extraction of implicit aspects. This paper provides a survey on recently proposed techniques for detecting implicit aspects. We have classified the studies according to approaches they have followed, also specified limitations and future work stated by authors. We have discussed different issues in implicit aspect extraction which will give directions for future research.

1. . Introduction

Sentiment analysis (or opinion mining), is a field of research which deals with the analysis of user's opinions, sentimentsexpressed in written text. SA is currently very dynamic research area due to the fast growth of internet and users' active participation for sharing, commenting and discussing over blogs, forums, social sites and shopping portals. SA can be helpful for manufactures, governments, businesses to get the feedback /impact of their product, service or decision. Sentiment Analysis is done at document-level, sentence level, and aspect-level. In document-level

sentiment analysis, the opinion words from the whole text are extracted and their polarity is detected and then polarity of whole document is determined. In Sentence-level SA, a single sentence is considered as small document and checked for subjectivity.

2. Aspect Based Sentiment Analysis

Aspect level Sentiment Analysis is a detailed analysis where opinions about various aspects of the entity are extracted. For example, in the review: "I am using this Phone since last month. It's superb. I totally love this phone; the LED display is beautiful.",phone is the entity and aspect is LED display. Superb, Beautiful are the sentiments towards the entity and its aspect.

Normally people tend to express opinions on some parts, aspects of a product or service not as a whole, in this case document level SA may in insufficient. Aspect-level sentiment analysis can give more useful knowledge about the author's opinion on various aspects of product or service. In some cases, sentence level SA may not detect polarity correctly. For example, a sentence like "I liked the display, but the battery life is only up to one day" contains aspects display and battery life. If we detect polarity at sentence level, it would result in a neutral polarity value because the sentimentsexpressed are of opposite polarity. So, extraction of aspects is essential to extract opinion-target pair.

3. Subtasks in Aspect Based Sentiment Analysis

Aspect/feature-level SA is done in phases as shown in fig 1. Extraction of an entity, its aspects and opinion word are done in aspect and opinion extraction phase. In second phase, i.e. sentiment lexicon phase, the polarity of extracted opinion words is determined. Finally, opinion summarization is done from all aspects and opinion word polarity.
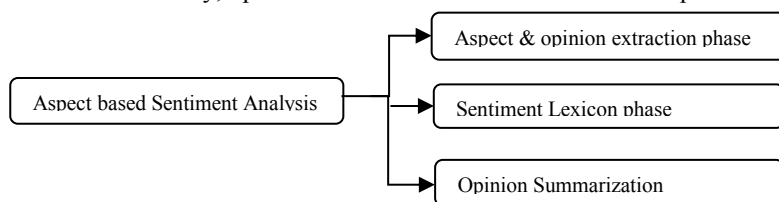

Fig.1 Phases in ABSA

In aspect extraction the aspects are identified from the opinionated text. These aspects can beexplicit and implicit aspects. The terms that are explicitly present in opinionated sentence are explicit aspects. Consider this review: "it's very light -weight and we can get amazing pix too", the aspect weight is specified explicitly. But in the sentence "It is very light. You can carry it everywhere", the aspect weight is implicit.

One of the recent surveys on ABSA is by Schouten et. al. [1]. In their survey, approaches for aspect detection, sentiment analysis and joint methods for aspect detection and sentiment analysis are specified and grouped them according to the type of algorithm used. The issues in detecting sentiments when comparative opinions, conditional sentences, negations and other modifiers are present, aggregating and presenting sentiment scores are discussed.

In [2] Rana et. Al. has given detailed summary of various aspect extraction techniques and approaches. They categorized the techniques according to the adopted approach and comparative analysis is given for these approaches of aspect extraction. Their study focused more on explicit aspects as there are numerous techniques available for explicit aspect extraction. They specified some approaches for implicit aspect extraction.

Extraction of explicit aspect has been extensively studied and many approaches are proposed for it. But, inadequate work has been done for extraction of implicit aspects. Extraction of implicit aspects is challenging but also important as implicit aspects are often found in opinionated document. In this survey we focused techniques for implicit aspect extraction. We surveyed some of the published literature on implicit aspect extraction since 2014 as work done before 2014 has been discussed in [2].

4. Approaches for Implicit Aspect Extraction

Researchers used many different approaches for implicit aspect extractions like Supervised, unsupervised semi supervised, hybrid. Compared to unsupervised and semi supervised, supervised techniques are mostly used in wide

variety of domains. Table 1 gives summary of all the approaches in chronological order, and we also specified limitations and future direction stated by the authors. For getting more understanding, we further classified the work depending on techniques used for aspect detection.

Table 1. Approaches for Implicit Aspect Extraction

Paper	Approach	Domain	Language	Limitations	Future Work
Sun et al (2014) [3]	Context Based	23656 mobile phone users review	Chinese	-Cannot figure out implicit features in small scale corpora -Considered only adjectives as opinion words	-Focus on detection of opinion words other than objectives -improve accuracy for the small-scale corpus
Poria et al. (2014) [4]	Unsupervised, rule-based approach	SemEval 2014 dataset	English	-Accuracy of method depends on thedependencyparseraccuracy and the opinion lexicon	-discover additional rules for extracting aspects and combine existing rules for complex aspect extraction
Schouten and Frasincar (2014) [5]	Supervised	Restaurant and productreviews	English	-For the datasets having less data but more unique implicit features the results are not that good to be useful in practice.	-learn different threshold for each implicit feature -extraction of more than one implicit feature.
Xu & Zang (2015) [6]	supervised, Svm based approach	cell phone review	Chinese	- The training set has sentences having only one explicit feature - the sentences containing explicit and implicit featuresboth areconsidered as explicit sentences. - sentences containing infrequent explicit and implicit aspects are ignored	- combining some novel classifiers and attribute selection methods for implicit feature detection - apply the proposed approach for other languages' implicit product review
Dosoula et.al. (2016) [7]	Supervised	Restaurants reviews	English	-performanceof the classifier depends on how the number of implicit features per sentence are distributed.	-Use advanced machine learning techniques, like SVM for training a classifier for multiple implicit features. -rule learning methods could be used for determining presence of multiple implicit features
Mohammed, B. (2016) [8]	Corpus based & dictionary-based model	products and restaurant datasets	English	-considered implicit aspect terms implied only by adjectives.	Use hybrid model to improve the underlying Naïve Bayes model and check for aspects in combination of adverbs, adjectives and verbs
Panchendraraj an et al (2016) [9]	Cooccurrence between opinion and other words	1000 restaurant reviews	English	-domain specific model	-improve the model, and extend to other domains by identifying relationships between domain specific aspects and modelling them as a hierarchy
Chatterji,S et al. (2017) [10]	supervised	Tweets, reviews and news	English	-the system may incorrectly tag the word if ambiguous opinion words which are used to describe different aspects in different domains are present	Extend the scope by including more product types
Bhatnagar V. et al (2016) [11]	Supervised	Tourism Domain	English	-domain specific Research -considering frequently occurring nouns as explicit aspects cannotcompletelychoose the correct aspects	-remove the different ambiguities present by considering the different contexts. while calculating the aspects - finding the aspects in reviews containing sarcasm and complex reviews.
Chen, H. Y., & Chen, H. H. (2016) [12]	corpus based	Hotel Review dataset from booking.com.	Chinese	-considering an implicit opinion and its neighbour explicit opinion possibly have the same aspect and polarity, neglected the case when aspect term or opinion word is missing in neighbouring sentence	-implicit aspect and polarity recognition for neglected cases.

Jinzhan Feng et. Al (2018) [13]	deep CNN&sequential algorithm	1.25 million mobile phone reviews	Chinese	- mainly focused on mobile phone reviews	-getting the appropriate model parameters
Zainuddin, et al [14]	Dependency Parsing	Hate crime review Tweets	English	-considered only subjective tweets	-experiment for social media data like YouTube and Facebook
Schouten, Kim, et al. [15]	Unsupervised Association rule mining	SemEval 2014Dataset	English	-unsupervised method requires tuning for multiple parameters	-inserting external knowledge to check improvement in results&dealingwith imbalanced data

4.1. Unsupervised rule-based approach

Poria et. al [4] demonstrated a technique for explicit and implicit aspect extraction from opinionated text. A rule-based approach using common-sense knowledge and sentence dependency trees is presented for detection of implicitas well as explicit aspects.Zainuddin, Selamat, and Ibrahim [14] used dependency parser to find grammatical relationship between aspects and opinions. These relationships can be useful for finding implicit aspects.

4.2. Frequency-based or statistical approach

A context-based method for implicit aspect extraction in Chinese product reviews is proposed by Sun et al. [3]. The implicit features extraction is done in three steps. First Co-occurrence matrix which shows relationship between product aspect and opinion words and is built using corpora training set. Check for presence of an implicit feature is done and possible implicit features set is built in second step. In third step, the scores of possible implicit features are calculated by checking opinion words and the implicit features context information and correct implicit feature is detected. This method mainly studies implicit feature extraction and not discuss ways for combining results and finding polarity.

A supervised approach is proposed by Schouten and Frasincar [5] for identification of implicit aspects. They used product and restaurant review dataset domain. The list of implicit aspects, list of unique lemmas and their respective frequencies is generated first and then score i.e. the ratio of co-occurrence of each word and frequency of the word for each implicit aspect is computed. The aspects having score greater than the defined threshold are identified.

The paper [7] by Dosoula et. al. is the extension of their previously proposed implicit feature algorithm. A classifier predicting the occurrence of multiple implicit features is built using a score function. The score function is based on number of nouns, adjectives, commas, and number of 'and' words. Using logistic regression, the function parameters are estimated, and a threshold set for performance improvement. The feature detection part of algorithm then checks for one or more implicit features using the prediction of above classifier.

4.3. Corpus based approach

Chen, H. Y., & Chen, H. H. [12] considered the problem of identifying aspects when both aspect and polarity are implicit. Under the observation that implicit opinion and its neighboring explicit opinion have the same aspect and sentiment, they built the implicit opinion corpus labeled with aspect class and polarity.

4.4. Pattern-based approach

Chatterji, S et al. [10] Stated that there is close association between explicit and implicit aspects. Hence, instead of considering explicit and implicit aspect separately, all these correlated aspects can be represented in a network-like structure. They proposed a framework (AspectFrameNet) for sentiment analysis. The system learns theaspect pattern in review text and the patterns of the aspects and the sentiments for these aspects are aggregated.

Schouten, Kim, et al. [15] proposed a unsupervised and supervised technique. In Unsupervised method association rule mining is applied on co-occurrence frequency data obtained from a corpus to find aspect categories. for each

aspect category a set of seed words are introduced and association rules of the form [notional word → category] are mined.

4.5. Hybrid approach

Mohammed, B. [8] proposed a hybrid approach that combines Wordnet dictionary based and corpus-based methods for extraction of implicit aspect terms. The approach works in phases named Implicit Aspect Representation, learning model enhancement and implicit aspect identification. In first phase for all adjectives from corpus, list of wordnet related words are extracted and then for each aspect from training data, all the relative adjectives are listed. In second phase using implicit aspect representation from phase 1, Naïve Bayes classifier is trained for detecting implicit aspects. And in final phase the naïve Bayes classification is tested for all terms with respect to aspects. The model is improved by eliminating the noisy WN related words. The approach is tested on two different web corpora.

4.6. Machine Learning based Approach

Xu and Zang [6] proposed a semi-supervised technique for detecting implicit aspects for Chinese reviews. Instead of using traditional feature selection method, LDA based explicit topic model is constructed. For each product feature a classifier is generated. Explicit sentences and their matching features are extracted from the review text and treated as training examples to build the classification models. Relevant explicit sentences are positive cases and relevant non-explicit sentences are negative cases for the training model of each feature. Then SVM classifiers are built and applied to distinguish the non-explicit sentences.

Panchendrarajan et. al [9] have considered the situation where an opinion word is used to describe different aspects. This paper presented a method to recognize multiple implicit aspects in a sentence. Separate models are created for identification of implicit and explicit aspects. These models used manually labelled training data set. First model uses maximum entropy classification technique for explicit aspect identification. Second model finds opinion words and their associated implicit aspects. Double propagation is used to extract opinion target. To improve the accuracy in presence of multiple interrelated aspects, entities and their aspects are modelled as a hierarchy. 1000 restaurant reviews are used as the training data set.

Bhatnagar V. et al [11] proposed framework and algorithm for identification of the explicit and implicit aspects in the tourism domain. The frequently occurring nouns are considered as explicit aspects. The supervised machine learning algorithm CRF and Stanford's NER is used for identifying implicit aspects. Here they considered frequently occurring nouns as aspect. As the same word may have different POS recognition, this consideration may not pick correct aspect.

Deep convolution neural network is combined with sequential algorithm by Jinzhan Feng et. al [13].   The feature vector built by aggregating words vectors, part of speech vectors, and dependent syntax vectors extracted from the words is given as input for training the deep convolution neural network, and sequential algorithm is then used for finding the sentiment expressed in the sentence.

5. Datasets for aspect-based sentiment analysis

Researchers used datasets from different domains for aspect extraction. Mostly used dataset for ABSA are electronic product review like mobile phone, computer reviews. Hotel and tourism reviews [5, 7, 9, 12] are also used by some researchers for ABSA. The reason may be there are vast number of aspects associated with cell phones like camera, appearance, interface, weight, performance, memory and for hotel like food, service, ambiance, cost, location. Most research work focus on English language and only few considered datasets in Chinese language [3, 12, 13]. It is observed that research in ABSA is domain specific. It is very difficult to generalize the ABSA method which will work for all domains as same aspect word can have different meaning in different domain.

6. Evaluation Measures used in ABSA

For evaluation of performance of ABSA model, the mostly used evaluation metrics are accuracy, precision, recall and F1-measure. Accuracy is the proportion of correctly predicted samples. But accuracy is insufficient measure if the dataset is unbalanced, so precision, recall and F1-measure comes into picture. The precision is proportion of correct predictions among all positive label samples and recall is proportion of correct predictions among all positive predictions. F1-measure is equally weighted average of recall and precision.

7. Issues in Implicit Aspect Extraction

Detection of implicit feature is challenging as people express their views very differently. The sentences may not be grammatically correct, abbreviations may be used, and language habit differs per person. The following subsections discuss issues in extraction of implicit aspects.

7.1. Multiple implicit features per sentence

When single implicit feature is present then we can easily associate sentiment of whole sentence as a sentiment of that feature but when multiple implicit features are present then finding sentiment of feature become more challenging. This is because the features might have different polarity than the overall polarity of sentence. For example, in the statement "Pictures taken can get blurred because of lack of image stabilizer but overall a great option for given budget", two different aspects camera quality and price are mentioned implicitly. Papers [7] & [9] focused on this issue.

7.2. Ambiguous opinion words

The opinion word can be used to describe different aspect in same or different domains. For example, "pizza is very small", the aspect size is associated with food item and in "rooms are very small" the aspect size is associated with entity room. R. Panchendrarajan et. Al [9] handled the issue when an opinion word can be associated with different aspects. For each opinion word a list of candidate aspects is extracted and using co-occurrence between the opinion word and other words score is computed. The aspect having highest score beyond threshold is chosen as implied aspect for that opinion word. The predicted implicit aspect is validated in last step. They also considered issue of multiple implicit aspects per sentence.

7.3. Rare word as aspect

In many cases the aspect or opinion word which is occurring less frequently is considered as noise. but it may convey important information. Paper [3] address this issue by considering not only frequency but the relationship between product feature and opinion word.

7.4. Noun or noun phrases can be implicit aspects

Most of research papers considered adjective and verbs as opinion word but in some cases noun or noun phrases can act as implicit aspect clues. For example, in a sentence, "overall it's a good deal", noun phase 'good deal' implies positive opinion. In paper [4] this issue is handled by exploiting common sense knowledge and sentence dependency tree. Using WordNet and SenticNet,an aspect knowledge base is developed for finding the aspect categories of IACs. The common-sense knowledge base comprises 30,000 multi-word expressions labelled by their polarity scores.

7.5. Double implicit problem

Opinion words act as clues in detecting implicit aspects. But when both aspect and opinion are implicit i.e. double implicit problem, finding aspect-sentiment pair becomes more challenging. Chen, H. Y., & Chen, H. H. [12] handled

this issue under the assumption that an implicit opinion and its neighboring explicit opinion possibly have the same aspect and polarity.

8. Conclusion

This paper surveys recently published work on implicit aspect extraction for sentiment analysis. We have classified the studies according to approaches they have followed, also specified limitations and future work stated by authors. Detecting implicit aspect is challenging but very important and more research must be done in this area.

We have discussed issues which can give directions for future research like; no state-of-the-art benchmark to evaluate extraction of implicit aspect, grammatically incorrect sentences, multiple implicit features per sentence, double implicit problem, noun or noun phrases as implicit aspects, presence of ambiguous opinion words i.e. word refer to different aspects in different domain.

Usually reviews are mix of explicit and implicit aspects, methods detecting implicit and explicit aspects jointly can be further useful. As the aspects are domain specific, the accuracy of aspect detection can be improved by inclusion of domain knowledge. Hybrid approach seems to be more appropriate for implicit aspect extraction. For some scenarios, Word based methods fail in detecting aspect and sentiment conveyed; there is need for semantic concept centric aspect level sentiment analysis.

References

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

[2] Rana, T.A. and Cheah, Y.N. (2016) "Aspect extraction in sentiment analysis: comparative analysis and survey" *Artificial Intelligence Review*, 46(4), pp.459-483.

[3] Sun, L., Li, S., Li, J. and Lv, J. (2014) "A novel context-based implicit feature extracting method" In *Data Science and Advanced Analytics (DSAA)*, 2014 International Conference on (pp. 420-424). IEEE.

[4] Poria, S., Cambria, E., Ku, L.W., Gui, C. and Gelbukh, A. (2014) "A rule-based approach to aspect extraction from product reviews" In *Proceedings of the second workshop on natural language processing for social media (SocialNLP)* (pp. 28-37).

[5] Schouten, K. and Frasincar, F. (2014) "Finding implicit features in consumer reviews for sentiment analysis" *International Conference on Web Engineering* (pp. 130-144). Springer, Cham.

[6] Xu, H., Zhang, F. and Wang, W. (2015) "Implicit feature identification in Chinese reviews using explicit topic mining model" *Knowledge-Based Systems*, 76, pp.166-175.

[7] Dosoula, N., Griep, R., den Ridder, R., Slangen, R., van Luijk, R., Schouten, K. and Frasincar, F.(2016) "Sentiment Analysis of Multiple Implicit Features per Sentence in Consumer Review Data" In *DB&IS (Selected Papers)* (pp. 241-254).

[8] Mohammed, B. (2016) "Hybrid approach to extract adjectives for implicit aspect identification in opinion mining" *Intelligent Systems: Theories and Applications (SITA)*, 2016 11th International Conference on (pp. 1-5). IEEE.

[9] Panchendrarajan, R., Ahamed, N., Murugaiah, B., Sivakumar, P., Ranathunga, S. and Pemasiri, A. (2016.) "Implicit Aspect Detection in Restaurant Reviews using Cooccurence of Words" *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 128-136).

[10] Chatterji, S., Varshney, N. and Rahul, R.K. (2017) "AspectFrameNet: a frameNet extension for analysis of sentiments around product aspects" *The Journal of Supercomputing*, 73(3), pp.961-972.

[11] Bhatnagar, V., Goyal, M. and Hussain, M.A. (2016) "A Proposed framework for improved identification of implicit aspects in tourism domain using supervised learning technique" *Proceedings of the International Conference on Advances in Information Communication Technology & Computing* (p. 56). ACM.

[12] Chen, H.Y. and Chen, H.H. (2016) "Implicit polarity and implicit aspect recognition in opinion mining" *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (Volume 2: Short Papers) (Vol. 2, pp. 20-25).

[13] Feng, J., Cai, S. and Ma, X. (2018) "Enhanced sentiment labeling and implicit aspect identification by integration of deep convolution neural network and sequential algorithm" *Cluster Computing*, pp.1-19.

[14] Zainuddin, Nurulhuda, Ali Selamat, and Roliana Ibrahim. "Hybrid sentiment classification on twitter aspect-based sentiment analysis." *Applied Intelligence* (2018): 1-15

[15] Schouten, Kim, et al. "Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data." *IEEE transactions on cybernetics* 48.4 (2018): 1263-1275.