Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 59 (2015) 244 – 253

Procedia Computer Science

International Conference on Computer Science and Computational Intelligence (ICCSCI 2015)

Aspect Extraction in Customer Reviews Using Syntactic Pattern

Warih Maharani, Dwi H. Widyantoro, Masayu Leylia Khodra *

Bandung Institute of Technology, Bandung, Indonesia

Abstract

Aspect extraction is an important step in opinion mining, to generate list of object, aspect and its opinions. Therefore, previous studies still give an opportunity to find the pattern of an aspect and opportunities in terms of aspect extraction performance. This paper propose syntactic pattern based on features observation to extract aspects from unstructured review, accompanied with a comprehensive analysis of varied pattern. This paper also provides some technical issues that arise based on performance evaluation and analysis using syntactic pattern extraction. The experimental results showed that the syntactic pattern approach had several weaknesses that need to be improved.

Keywords: aspect extraction; opinion mining; syntactic pattern

1. Introduction

With the explosive growth of social media, people can share their opinions easily in blogs, microblogs, comments, forum discussions and social network sites. Nowadays, people used to find user review and discussions in forums or e-commerce websites before buying a product [9]. However, due to the number of user reviews, it is difficult for customer to find a proper review in accordance to user needs. It encourages research on opinion mining and summarization [1, 25].

Generally there are two approaches in opinion mining, the traditional approach and the aspect-based approach [23]. The traditional approaches do not focus on an object or aspect discussed in the review document. While aspect-based

* Corresponding author. Telp. (022) 7564108
E-mail address: wmaharani@telkomuniversity.ac.id

doi:10.1016/j.procs.2015.07.545

approach focuses on the opinions of an aspect. Some research defines aspects as object features or product feature [23,24]. Aspect-based opinion mining consists of 2 main stages : 1) extraction of aspects and opinions, and 2) polarity classification [23]. The performance of polarity classification depends on the results of aspect extraction. So that the aspect extraction plays an important role to generate a complete list of objects, aspect and its opinions.

Turney used a certain set pattern to extract aspects and classify reviews into recommended and not recommended [25]. Popescu implemented OPINE which extract aspects using unsupervised learning [18,19]. A similar aspect extraction approach is used in Htay [19,25–27]. The results showed that there are opportunities to develop pattern-based extraction. The purpose of this research is to extract aspect with syntactic pattern-based approach. The contributions of this research are a comprehensive analysis of the pattern that has been used in previous studies with the proposed method using the new syntactic pattern based on features observations, with various pattern combinations.

The paper is organized as follows : Section 2 presents the related works of this paper ; Section 3 describes the aspect extraction ; Section 4 discusses the experiments ; Section 5 presents the discussion of the result of aspect extraction ; finally, Section 6 concludes the paper and challenges as well as opportunities for further research.

2. Related Works

Aspect-based opinion mining involved objects, aspects and followed by opinions [9]. Datasets commonly used in this research area such as electronic product reviews [2,18,21], movie reviews [3,8], reviews on hotel services, restaurant and travel agency [4], as well as specifics topics such as telephone conversations [5] or blogs [6,7].

The purpose of aspect extraction is to extract opinion sentences which contain one or more aspects and opinions. Several research used natural language processing techniques to extract objects, aspects and opinions [8,21]. These approach used features such as bag-of-words, term frequency (number of occurrences of the term in the document), term co-occurance (a feature that appears together, such as unigram/n-gram), POS (*Part of Speech*), opinion term (words that express positive or negative opinion), negation and syntactic dependency [9].

In general, Part-of-speech (POS) and syntactic parsing are used to extract noun phrases [10–13,18,21,24–27]. Turney used pattern-based approach to extract phrases and implemented Pointwise Mutual Information (PMI) to classify opinion based on the word "excellent" and "poor" [25]. Semantic orientation of the documents predicted based on the average semantic orientation of the adjective phrases and adverb phrases. Riloff proposed bootstrapping algorithm to classify subjective noun [13]. The subjective nouns can be used to extract aspects and opinions using subjectivity clues [13]. Popescu proposed unsupervised technique to extract aspect, which used syntactic dependencies produced by MINIPAR parser [18,19]. A similar approach is used in Kim and Hovy [14] .

Htay developed aspect extraction through opinion seedset and extraction patterns [26,27]. They identified pattern of opinion phrases through adjective, adverb, verb and noun [26,27]. While Zhang identified noun phrase feature to implied opinions. Zhang proposed candidate identification using statistical test and lexicon-based approach [15,22]. Existing research commonly used POS tagging and parsing techniques because they are relative simple and produce relatively high accuracy [21,24]. However, this technique can not extract aspects that are not explicitly mentioned in the datasets [21].

Supervised learning has shown better performance in aspects extraction [16,17]. Kobayashi proposed an aspect extraction method using a triple of <Subject, Attribute, Value> corresponding to product, aspect and polarity. They used machine learning techniques for this task [16,17]. One of the main challenges in supervised learning is the lack of training datasets with a limited to a particular domain. Instead unsupervised learning approach has relatively low performance, but this approach is able to handle variations and relatively extensive review domain [18,19]. These limitations can be overcome with a lexicon dictionary that contained semantic orientation value [20].

3. Aspect Extraction

The focus of this research is aspect extraction by comparing some patterns that represent an aspect. In addition to applying the pattern that has been used in previous studies, this research also extract new patterns based on the observation of features pattern and occurrence in the dataset. Based on the comparison of several extraction patterns, we can analyzed the weakness of the aspect extraction using syntactic pattern that can be developed further.

3.1. Data Analysis

This research used review datasets from Hu and Liu research [21,24,29]. Hu used dataset of some electronic products. Dataset used in this experiment are Canon S100, Canon SD500, Canon G3, Jukebox and DVD player Apex AD2600 Progressive-scan DVD player [17]. The dataset collected from Amazon.com website and Cnet.com. Some examples of dataset are as follows [20] :

a) memory[-1][s],battery[-1]##A larger memory card and extra battery are good things to buy.
b) support[-3][u]##apex does n't answer the phone .
c) player[-3][p]##after about an hour it starts skipping like the dvd 's are dirty .
d) light[+3][u],tiny[+3][u]##This camera is so light and tiny that I literally carry it everywhere
e) small[+1]##I want to start off saying that this camera is small for a reason.
f) small[+3],durable[2]## it is so small and it is so durable

"Memory [-1] [s], battery [-1]" are examples of aspect that explicitly appear in the dataset with its negative values. While the "support [-3] [u]" is an aspect that does not appear explicitly in the dataset. In addition there are some aspect that are not explicitly shown in the dataset, but it should be done by pronoun resolution, such as : player [-3] [p], the [-3] [p], dvd-r [+1] [p], video output [-3 [p]. Apparently there are some features that are labeled [u], but the word is seen explicitly in the sentence review, as in the example light [+3] [u], tiny [+3] [u] and screen [-3] [u]. Sentences (e) and (f) have a label small [+1], small [+3], durable [2], but we observed that these aspect can not be assumed as an aspect, but more likely to be the value of an aspect. So based on the observation of the entire dataset, aspect are grouped by aspect type :

a) datasets with complete aspect include explicit and implicit aspects
b) dataset with explicit aspect (without the [p] and [u] symbol)

Table 1. Data Analysis

Dataset	a	b	c
Canon S100	300	182	17
Canon SD500	229	148	0
Apex DVD Player	839	348	83

a : number of review sentences
b : number of aspect
c : number of aspect contain [p][u]

3.2. Preprocessing

In the early stages, POS tagging conducted to identify the noun phrase that will become aspects candidates by using the Stanford-POS tagger[1]. This is because in general an aspect represented in the noun phrase [22,23]. We extract POStag of current word, POStag of a word before current word and POStag of a word after current word. In this research, we tokenized sentences into token, bigram and trigram, to represent feature for pattern extraction. We extracted these features from all sentence in Hu & Liu dataset [24].

This paper extracts aspects by applying a pattern-based approach that is taken from previous studies as well as to modify the new pattern. There are 3 sets of patterns are applied to aspect extraction, the pattern from Turney [25], the pattern from Htay [26] and new patterns based on features observations. Table 2 shows the Turney pattern (pattern A) and Table 3 shows the Htay pattern (pattern B).

Table 2. Pattern A [25]

No.	First word	Second word	Third word
1.	JJ	NN atau NNS	-
2.	RB/RBR/RBS	JJ	Not NN/NNS

3.	JJ	JJ	Not NN/NNS
4.	NN/NNS	JJ	Not NN/NNS
5.	RB/RBR/RBS	VB/VBD/VBN/ VBG	-

Table 3. Pattern B [26]

No.	First word	Second word	Third word
1.	JJ	NN/NNS	-
2.	JJ	NN/NNS	NN/NNS
3.	RB/RBR/RBS	JJ	-
4.	RB/RBR/RBS	JJ/RB/RBR/RBS	NN/NNS
5.	RB/RBR/RBS	VBN/VBD	-
6.	RB/RBR/RBS	RB/RBR/RBS	JJ
7.	VBN/VBD	NN/NNS	-
8.	VBN/VBD	RB/RBR/RBS	-

Based on the above two patterns, there are some patterns that are similar, such as patterns of numbers 1,2 and 5 from pattern A with the pattern number 1-5 from pattern B. In addition, there are several patterns from pattern B are similar, such as the pattern 1-2 and patterns 3 - 4.

3.3. Proposed Pattern

New pattern is obtained by features observations using bag-of-words feature, POStag pattern based on current POStag, (current-1) POStag and (current+1) POStag. Modifications made based on syntactic patterns that appear in these features. Based on these features observations and frequency that appears on the dataset, the pattern with a certain frequency threshold, will be adopted into the aspect extraction patterns. Based on the patterns analysis, there are some patterns that can represent an explicit object features as shown in Table 4. This approach has disadvantage which is not able to extract implicit product features. It is due to the implicit meaning contained in the phrase which can be determined by using natural language processing.

Table 4 shows the pattern C which is applied to the dataset as a new candidate aspect extraction.

Table 4. Pattern Extraction C

No.	First word	Second word	Third word
1.	NN/NNS	-	-
2.	VB	-	-
3.	DT/JJ/NN	NN/NNS	-
4.	NN/NNS/RB/RBR/RBS	JJ/VBN/VBD	-
5.	VBN/VBD	NN/NNS/RB/RBR/RBS	-
6.	JJ	NN/NNS	NN/NNS
7.	RB/RBR/RBS	JJ/RB/RBR/RBS	NN/NNS/JJ
8.	JJ	VB/VBN/VBD	NN/NNS
9.	NN/NNS	VB/IN/NN/NNS	NN/NNS
10.	NN/NNS	IN	DT – NN/NNS
11.	NN	TO	NN/NNS – NN/NNS

Pattern C has a different set of patterns with Turney and Htay [26,27]. These patterns obtained from features observations and frequency patterns that extracted from datasets. All these patterns have the same goal which is to identify and extract phrases that would be candidates product features. Another difference is the presence of a more comprehensive comparative analysis of the characteristics of different datasets were applied to three different patterns and its combinations.

4. Experiments

Experiments were performed by applying a combination of three different sets of patterns, a combination of Pattern A, Pattern B and Pattern C with several experimental scenarios :

1) The experiments using the entire dataset, includes explicit and implicit aspects
2) The experiment using a dataset consist of explicit aspects (not containing the [u] and [p] symbol)
3) The experiments using datasets that have been annotated and eliminate a portion of aspects containing [u] and [p] symbol

4.1. Experiment Scenario 1

The purpose of first experiment was to extract the aspects based on the pattern A, pattern B, pattern C and their combinations using a dataset that consist of explicit and implicit aspect. Table 5 shows the results of aspect extraction based on scenario 1.

Table 5. The result of scenario 1

Dataset	Precision	Recall	F-Measure
Pattern A			
Canon S100	0,421	0,611	0,498
Canon SD500	0,439	0,633	0,518
Apex DVD	0,386	0,431	0,407
Pattern B			
Canon S100	0,438	0,625	0,515
Canon SD500	0,471	0,641	0,543
Apex DVD	0,395	0,444	0,418
Pattern C			
Canon S100	0,489	0,645	0,556
Canon SD500	0,490	0,655	0,560
Apex DVD	0,398	0,465	0,428
Pattern A + C			
Canon S100	0,494	0,656	0,563
Canon SD500	0,501	0,664	0,571
Apex DVD	0,412	0,511	0,456
Pattern B + C			
Canon S100	0,501	0,674	0,574
Canon SD500	0,531	0,689	0,599
Apex DVD	0,425	0,524	0,469
Pattern A + B + C			
Canon S100	0,524	0,686	0,600
Canon SD500	0,545	0,692	0,612
Apex DVD	0,441	0,537	0,489

Based on the results analysis from the experiments above, there are some implicit aspects contained in the dataset, as indicated on the composition of datasets in the table in Section 3.1. The number of implicit aspects in the dataset Canon100 reach 25% of the total dataset, while CanonSD500 have 8.7% and Apex have more than 50% of implicit aspects.

The result showed that pattern combination produces a slight increase in Precision and Recall. It is due to the presence of a complementary pattern compared with the pattern set without combination.

4.2. Experiment Scenario 2

This experiment would like to analyze aspects that explicitly contained in the review. In this research, dataset was separated between explicit and implicit aspects, by eliminating aspects that contain [u] and [p] symbol. Table 6 shows the experiments results that contains explicit aspects (without the [u] and [p] symbol).

Table 6. The result of scenario 2

Dataset	Precision	Recall	F-Measure
Pattern A			
Canon S100	0,430	0,618	0,507
Canon SD500	0,439	0,633	0,518
Apex DVD	0,390	0,441	0,413
Pattern B			
Canon S100	0,440	0,629	0,517
Canon SD500	0,471	0,641	0,543
Apex DVD	0,394	0,454	0,421
Pattern C			
Canon S100	0,490	0,651	0,559
Canon SD500	0,490	0,655	0,560
Apex DVD	0,402	0,472	0,434
Pattern A + C			
Canon S100	0,497	0,660	0,567
Canon SD500	0,501	0,664	0,571
Apex DVD	0,423	0,512	0,463
Pattern B + C			
Canon S100	0,508	0,681	0,581
Canon SD500	0,531	0,689	0,599
Apex DVD	0,428	0,542	0,478
Pattern A + B + C			
Canon S100	0,540	0,690	0,605
Canon SD500	0,550	0,692	0,612
Apex DVD	0,452	0,541	0,492

The second experiment showed that there is no significant increase in Precision and Recall by eliminating the implicit aspects that are marked with the [u] and [p] symbol. By using this extraction pattern, certainly not able to extract the implicit aspects, so we need further development to extract the implicit aspects.

In addition to the implicit aspects, there are some aspects that are not expressed in Noun Phrase, such *as "light", "tiny", "work", "small",* as shown in the following sentence:

"This camera is so ***light*** *and* ***tiny*** *that I literally carry it everywhere with me, which is great considering I'm an artist and like to capture interesting items."*

"I've often struggled to get something to ***work*** *the way I want it, and hate to have to carry the manual with me all the time!"*

"It's so ***small*** *and light, even with a spare battery, you really can take it anywhere."*

The number of aspects that are not in Noun phrase form approximately 33 features for data Canon S100, 17 features for data Canon SD500 and more than 100 features for data Apex DVD. So for the case above, it is necessary to re-annotate the dataset and analyze the specific pattern extraction on the explicit aspects.

4.3. Experiment Scenario 3

The goal of third experiment is to extract the aspects based on the pattern A, pattern B, pattern C and its combinations by using several datasets that have been re-annotation and eliminates aspects containing [u] and [p] symbol. Some examples of re-annotated datasets are:

small[+1]##I want to start off saying that this camera is small for a reason.
small[+3],durable[2]## it is so small and it is so durable
fun[+3]##I bought this little guy a few weeks back, and I have to saythat I never had so much fun with a new toy as this.
play[+2], dvd-r[+1]##what got me to buy was the reviewer that said it would play dvd-rs fill of files (e. , mp3s)
.

Table 7. The result of scenario 3

Dataset	Precision	Recall	F-Measure
Pattern A			
Canon S100	0,541	0,713	0,615
Canon SD500	0,532	0,701	0,604
Apex DVD	0,445	0,534	0,485
Pattern B			
Canon S100	0,546	0,763	0,636
Canon SD500	0,556	0,742	0,635
Apex DVD	0,476	0,582	0,523
Pattern C			
Canon S100	0,591	0,784	0,673
Canon SD500	0,680	0,763	0,719
Apex DVD	0,501	0,546	0,522
Pattern A + C			
Canon S100	0,594	0,797	0,680
Canon SD500	0,689	0,780	0,731
Apex DVD	0,521	0,578	0,548
Pattern B + C			
Canon S100	0,604	0,799	0,687
Canon SD500	0,693	0,794	0,740
Apex DVD	0,532	0,574	0,552
Pattern A + B + C			
Canon S100	0,625	0,789	0,697
Canon SD500	0,699	0,805	0,748
Apex DVD	0,554	0,590	0,571

The third experiment showed that the separation between explicit and implicit aspects lead to an increase in Precision and Recall. It shown in Table 7. However, the system has not been able to extract all of existing explicit aspects. For example, there are several aspects that are not expressed in Noun Phrase, such as "use", "looks", "ease of use" that contained in the following sentences :

*"Excellent picture quality and so simple to **use**!!!"*
*"**Looks** great."*
"I was looking for something that wasn't too complicated to use and to help the pics I post on Ebay show more detail."

In addition there are several phrases that also extracted, but it is not an aspect. Table 8 shows the phrases that are also extracted, but these are not aspects. It called false positives in aspect extraction.

Table 8. False Positive Phrases

Pattern	Phrases
JJ – NN	Digital camera, Digital photography, Digital zoom, Digital image, Digital equipment
NN – JJ	Person bright, zipper extra, body popular, lens little, situation quiet, dark nice, card extra, look standard, idea front, pounch best, rendering perfect
RB – JJ	Less warm, far superior, actually better
RB – VB	First saw, already bought, quickly showed
VB – NN	Bought camera, saw picture

As shown in Table 8, examples of phrases that are often extracted, but not aspects are phrases consist of JJ - NN, such as "digital camera", "digital zoom" or "digital image" that are extracted as an aspect. In addition there are another pattern that often extracted, such as patterns RB - JJ, RB - VB and VB - NN as shown in table 8 above. Based on the experimental results, it need a post-processing to filter the candidate aspects, such as Bross research [33], so the Precision and Recall can be increased.

5. Discussion

The aim of this research is to explore the strengths and weaknesses of a pattern-based approach, and discover new patterns with some combination of existing pattern with the new pattern. So based on the experimental results, there are still some problems to be solved, including explisit and implicit aspect extraction. This paper focus to extract only explicit aspects. So that in future research need to apply an approach that can extract implicit aspect, for example by applying natural language processing such as coreference and pronoun resolution and identification of implicit aspect, so that can produced a complete list of aspect include explicit and implicit [28,33]. Another alternative approach is to apply a knowledge-based approach, dependency relation or corpus-based [29–31,33].

Another issues are based on the pattern that had been applied, there are some noun phrase that also extracted, but not an aspect. As an example sentence review :

*"I researched **digital camera** for a month before purchasing the S100."*
*"This is the best camera you can have for **digital images** if you want to"*

Based on the "JJ – NN" pattern, hence the phrase "digital camera" and "digital images" will be extracted, but it was not an aspect. We apply corpus-based approach to overcome this problem. The implementation of the post-processing provides performance improvement. In addition, the pattern that has been implemented is not able to identify aspects and opinion pairs separated by a few tags. For example :
"The picture quality of this camera is amazing"

The system should be able to extract the object "camera" that has an aspect "picture quality" with the value "amazing" that has positive orientation. But this study did not focus on finding a solution to solve these problems because the purpose of this study was to conduct a comprehensive analysis of pattern-based aspect extraction. There are several alternative solutions that can be used to solve these problems, such as using dependency parser or semantic-based approach [32–34].

6. Conclusion and Future Works

Aspect extraction is an important stage in the aspect-based opinion mining. This research focuses on the syntactic-based approaches by comparing several pattern-based extraction. Pattern-based aspect extraction can only extract aspect that explicitly mentioned in dataset, but can not be used to extract the implicit aspects. Implicit aspect can not be extracted with existing these syntactic patterns, but requires a semantic-based approach. While variations in aspects either explicit or implicit aspects pretty much met in real conditions, thus requiring a more representative approach as by adding post-processing stage, or using natural language processing.

References

1. Pang B, Lee L, Vaithyanathan S. Thumbs up?: sentiment classification using machine learning techniques. Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics; 2002. p. 79–86.
2. Liu B, Hu M, Cheng J. Opinion observer: analyzing and comparing opinions on the web. Proceedings of the 14th international conference on World Wide Web. ACM; 2005. p. 342–351.
3. Kennedy A, Inkpen D. Sentiment classification of movie reviews using contextual valence shifters. Computational Intelligence. Wiley Online Library; 2006;22:110–125.
4. Zhang Z, Ye Q, Zhang Z, Li Y. Sentiment classification of Internet restaurant reviews written in Cantonese. Expert Systems with Applications. Elsevier; 2011;38:7674–7682.
5. Wang D, Liu Y. A Pilot Study of Opinion Summarization in Conversations. ACL. 2011. p. 331–339.
6. Mishne G, Balog K, Rijke M de, Ernsting B. Moodviews: Tracking and searching mood-annotated blog posts. Proceedings International Conf. on Weblogs and Social Media (ICWSM-2007). 2007. p. 323–324.
7. Ku L-W, Liang Y-T, Chen H-H. Opinion Extraction, Summarization and Tracking in News and Blog Corpora. AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs. 2006. p. 100–107.
8. Zhuang L, Jing F, Zhu X-Y. Movie review mining and summarization. Proceedings of the 15th ACM international conference on Information and knowledge management. ACM; 2006. p. 43–50.
9. Liu B. Sentiment analysis and opinion mining: synthesis lectures on human language technologies. Morgan & Claypool Publishers. 2012;
10. Zhai Y, Chen Y, Hu X, Li P, Wu X. Extracting Opinion Features in Sentiment Patterns. Information Networking and Automation (ICINA), 2010 International Conference on. IEEE; 2010. p. 1–115.
11. Luo Y, Huang W. Product Review Information Extraction Based on Adjective Opinion Words. Computational Sciences and Optimization (CSO), 2011 Fourth International Joint Conference on. IEEE; 2011. p. 1309–1313.
12. Turney P. Learning to extract keyphrases from text. 1999;
13. Riloff E, Wiebe J, Wilson T. Learning subjective nouns using extraction pattern bootstrapping. Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4. Association for Computational Linguistics; 2003. p. 25–32.
14. Kim S-M, Hovy E. Determining the sentiment of opinions. Proceedings of the 20th international conference on Computational Linguistics. Association for Computational Linguistics; 2004. p. 1367.
15. Zhang L, Liu B, Lim SH, O'Brien-Strain E. Extracting and ranking product features in opinion documents. Proceedings of the 23rd international conference on computational linguistics: Posters. Association for Computational Linguistics; 2010. p. 1462–1470.
16. Kobayashi N, Inui K, Matsumoto Y. Opinion mining from web documents: Extraction and structurization. Information and Media Technologies. Information and Media Technologies; 2007;2:326–337.
17. Kobayashi N, Inui K, Matsumoto Y, Tateishi K, Fukushima T. Collecting evaluative expressions for opinion extraction. Natural Language Processing–IJCNLP 2004. Springer; 2005. p. 596–605.
18. Popescu A-M, Nguyen B, Etzioni O. OPINE: Extracting product features and opinions from reviews. Proceedings of HLT/EMNLP on interactive demonstrations. Association for Computational Linguistics; 2005. p. 32–33.
19. Popescu A-M, Etzioni O. Extracting product features and opinions from reviews. Natural language processing and text mining. Springer; 2007. p. 9–28.
20. Voll K, Taboada M. Not all words are created equal: Extracting semantic orientation as a function of adjective relevance. AI 2007: Advances in Artificial Intelligence. Springer; 2007. p. 337–346.
21. Hu M, Liu B. Mining and summarizing customer reviews. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM; 2004. p. 168–177.
22. Zhang L, Liu B. Identifying noun product features that imply opinions. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2. Association for Computational Linguistics; 2011. p. 575–580.
23. Liu B. Sentiment analysis and subjectivity. Handbook of natural language processing. Chapman & Hall; 2010;2:627–666.
24. Hu M, Liu B. Opinion extraction and summarization on the web. AAAI. 2006. p. 1621–1624.
25. Turney PD. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics; 2002. p. 417–424.
26. Htay SS, Lynn KT. Extracting Product Features and Opinion Words Using Pattern Knowledge in Customer Reviews. The Scientific World Journal. Hindawi Publishing Corporation; 2013;2013.
27. Htay SS. Biologically Inspired Opinion Mining: Features Extraction based on Linguistic Patterns for Customer.
28. Broß J ürgen, Ehrig H. Terminology Extraction Approaches for Product Aspect Detection in Customer Reviews. CoNLL-2013. Citeseer; 2013;222.
29. Ding X, Liu B, Yu PS. A holistic lexicon-based approach to opinion mining. Proceedings of the 2008 International Conference on Web Search and Data Mining. ACM; 2008. p. 231–240.
30. Wu Y, Zhang Q, Huang X, Wu L. Phrase dependency parsing for opinion mining. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3. Association for Computational Linguistics; 2009. p. 1533–1541.
31. Lu Y, Castellanos M, Dayal U, Zhai C. Automatic construction of a context-aware sentiment lexicon: an optimization approach. Proceedings of the 20th international conference on World wide web. ACM; 2011. p. 347–356.

32. Gindl S, Weichselbraun A, Scharl A. Rule-based opinion target and aspect extraction to acquire affective knowledge. Proceedings of the 22nd international conference on World Wide Web companion. International World Wide Web Conferences Steering Committee; 2013. p. 557–564.
33. Bross J, Ehrig H. Automatic construction of domain and aspect specific sentiment lexicons for customer review mining. Proceedings of the 22nd ACM international conference on Conference on information & knowledge management. ACM; 2013. p. 1077–1086.
34. Weichselbraun A, Gindl S, Scharl A. Extracting and grounding context-aware sentiment lexicons. IEEE Intelligent Systems. 2013;28:39–46.