Aspect-Based Sentiment Analysis Using Grammatical Rules, Word Similarity and SentiCircle

Ari Firmanto[1]   Riyanarto Sarno[1*]

[1]*Department of Informatics*,
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
* Corresponding author's Email: riyanarto@if.its.ac.id

Abstract: Product reviews can be used as suggestion for companies to improve their services in this digitalization era. These reviews can be presented in more detail using aspect based sentiment analysis. In this research, aspect-based sentiment analysis method is proposed using grammatical rules, word similarity, and SentiCircle. This method began with extracting the candidate aspects rules based on phrase detection in constituency parse. The candidate aspects were categorized using word similarity. Word similarity calculated the similarity value between the candidate aspects and the keywords obtained from Wikipedia. To determine sentiment polarity, SentiCircle is used. SentiCircle can capture the contextual sentiment from the data. The result showed that the proposed method was able to categorize aspects correctly, with the highest f1-measure value of 84%, while sentiment analysis produced the highest f1-measure of 87%.

Keywords: Grammatical rule, SentiCircle, Aspect categorization, Sentiment analysis, Word similarity, Aspect-based sentiment analysis.

1. Introduction

In this digital era, information can be easily obtained. One such information is product reviews. Product reviews are written by users to provide recommendations to companies. Thus, it is very important to analyze this review. One technique that can analyze product reviews is sentiment analysis [1]. However, sentiment analysis provides general social sentiment, so a method that can provide a more detailed recommendation based on certain aspects is needed. Such a method is called aspect based sentiment analysis [2].

Previous research on sentiment analysis was applied in analyzing reviews in various fields, including hotels, films, online shopping, and restaurants. Suhariyanto et al. [3] analyze the sentiment of the movie reviews on RottenTomatoes using SentiWordnet. Bagus et al. [4] analyze the sentiment of the product reviews on the Amazon by addressing local and global contexts. As well as,

Dewi et al. and Reza et al. [5, 6] conduct an experiment on aspect based sentiment analysis on hotel reviews using topic modelling and machine learning methods.

In aspect-based sentiment analysis, it is generally divided into three sub-processes. The processes are aspect extraction, aspect aggregation, and sentiment analysis [7]. Aspect extraction is the process to extract the candidate aspects from reviews. These candidate aspects can be explicitly written or implicitly contained in a sentence [8]. Aspect aggregation is the process to group the candidate aspects into pre-defined aspect [9]. Sentiment analysis is the process to assign the polarity of sentiment into aspects. Sometimes in the process to determine the sentiment, there are opinions that depend on the particular context. For instance, in the following reviews:

1. Price is high but the food is good, so I would come back again;

2. The takeout is great too since they give high quality tupperware as well;

In the first sentence, the word high has negative sentiment, but in the second sentence, it has a positive sentiment. Thus, it is very important to pay attention to the context in determining sentiment in an aspect.

Multiple aspects can be found in a sentence in the reviews. For instance, "service was divine, oysters where a sensual as they come, and the price can't be beat!!!". This review describes service and price of the restaurant. Reza et al. [6] in the previous research, assumes that in the sentence of reviews contains only one aspect. This may be acceptable in the simple sentences, but it can't in the sentences contain multiple clauses.

Furthermore, the rule-based on dependency relations are not effective in the long sentence [21]. For instance, "i tend to judge a sushi restaurant by its sea urchin, which was heavenly at sushi rose". Using dependency relation, aspect "sushi" and opinion "heavenly" cannot be extracted because there is no relation between them. Thus, it is very important to compose the rules that can extract both aspect and opinion in this kind of sentence. This paper proposes the rule method that can extract multiple aspects in the sentences and determine opinions based on the surrounding context. This rule starts with splitting the sentence into several clauses or fragments to facilitate analysis of sentence structure. Furthermore, for each clause, the phrase is determined using phrase extraction. From this phrase, the candidate aspects and opinion words are extracted. The next process is aspect aggregation. The candidate aspects obtained are categorized with the aspect keywords using word similarity. These keywords are generated using TF-ICF [10] algorithm from Wikipedia.

To determine the sentiment polarity based on the surrounding context, SentiCircle is used [11]. SentiCircle uses the hypothesis that words that appear in the same context tend to have the same meaning [12]. This process starts with determining term degree correlation between word and the context words, determining the prior sentiment using opinion lexicon, and calculating sentiment score represented in the polar coordinate system. In this research, SentiWordnet [13] is used as an opinion lexicon to determine the prior sentiment. Then, the word sense disambiguation method is applied to determine the right sense that can improve the performance of sentiment analysis [14-16.

SentiCircle is dependent on the completeness of the data to calculate word correlations. The more complete the dataset, the better the result obtained. In this research, context words expansion is proposed to complete the word in the dataset. This expansion is based on synonyms from the context of the word and co-occurrence of the word context with others in the dataset.

In the remainder of this paper, Section 2 gives the detailed related theory used in this research. Section 3 describes the detailed of proposed method in aspect categorization and sentiment analysis. The results of both proposed method are evaluated and discussed in Section 4. Finally, the researchers conclude with summary and some future research directions in Section 5.

2. Related theory

This section explains about several theories related to the research.

2.1 Keyword term list for aspect category

In this research, restaurant reviews from TripAdvisor are used as a dataset. This dataset had been annotated manually into 4 aspect categories. The aspect categories used in this research can be seen in Table 1, while Table 2 shows the Wikipedia and external links used in keyword extraction.

Table 1. Aspect category and definition

Aspect Category	Definition
Food	Opinions that focusing on the food in general or in terms of specific dishes, dining options etc.
Service	Opinion focusing on the service, on the promptness and quality of restaurant's service in general, the food preparation, the staff's attitude and professionalism, the wait time, and the options offered.
Ambience	Opinions that focusing on the atmosphere or the environment of the restaurant's interior or exterior space (terrace, yard, garden), the décor, and entertainment options.
Price	Opinions that focusing on value, price, and cost of the restaurant.

Table 2 Wikipedia links of aspect keywords

Aspect Category	Wikipedia Links	External Links
Food	https://en.wikipedia.org/wiki/Food https://en.wikipedia.org/wiki/Cuisine	http://www.foodtimeline.org Cookbook:Table of Contents
Service	https://en.wikipedia.org/wiki/Customer_service https://en.wikipedia.org/wiki/Waiting_staff	Swanson's Unwritten Rules https://www.bbc.com/news/magazine-12123463
Ambience	https://en.wikipedia.org/wiki/Theme_restaurant https://en.wikipedia.org/wiki/Spirit_of_place	https://www.elitetraveler.com/finest-dining/what-makes-a-good-restaurant-atmosphere http://restauranttallent.com/what-makes-a-good-restaurant-atmosphere/ Quebec Declaration on The Preservation of The Spirit of Place
Price	https://en.wikipedia.org/wiki/Price https://en.wikipedia.org/wiki/Value_(economics)	https://en.wikisource.org/wiki/1911_Encyclopædia_Britannica/Price http://libraryguides.missouri.edu/pricesandwages

2.2 Pre-processing

In general, pre-processing used in natural language processing is tokenization, normalization, remove punctuation, lemmatization, stop words removal, and spelling correction.

2.3 Word embedding + word similarity

In natural language, words can be similar lexically and semantically [17]. Words are similar lexically if they have a similar character sequence. Words are similar semantically if they have same meaning. Word embedding is the representation of word vectors that can capture the context of word in document. In this research, FastText [18] is used for word embedding. To calculate similarity between two words in word embedding, cosine similarity is used.

$$Similarity(w_i, w_j) = \frac{\sum_{m-1}^{K} w_i{}^m w_j{}^m}{\sqrt{\sum_{m-1}^{K} (w_i{}^m)^2} \sqrt{\sum_{m-1}^{K} (w_j{}^m)^2}}$$
(1)

In Eq. (1), similarity measures the distance between first word (w_i) and second word (w_j). $\sum_{m-1}^{K}$ is the number of iterations from m to K word. The result of similarity ranges from -1 to 1.

2.4 Keyword extraction (TF-ICF)

Term Frequency-Inverse Cluster Frequency (TF-ICF) [10] is the method of weighting terms based on information from document on cluster. In general, TF-ICF calculates term frequency from the document on each clusters. ICF value is calculated based on the number of clusters that exists and the number of clusters that contains term i (cf_i). The equation of ICF can be seen in Eq. (2)

$$ICF_i = 1 + \log \frac{c}{cf_i}$$
(2)

TF-ICF weighting is calculated using Eq. (3), where TF_{ji} is the term frequency of term i in cluster j and ICF_i is the ICF value of term i.

$$TF\text{-}ICF_i = TF_{ji} \times ICF_i$$
(3)

2.5 SentiCircle

SentiCircle [11] is a method to capture the contextual semantics and sentiments of words. The idea of this method is that sentiment of a term is not static, but rather depends on the context which the term uses. It follows the distributional hypothesis stating that words that occur in similar contexts tend to have similar meanings [12]. SentiCircle represents sentiment model using polar coordinate system. It assumes the following processes:

1. Generating term degree of correlation ($TDOC$) that represents the degree correlation between a term and context term.

$$TDOC(m, c_i) = f(c_i, m) \times \log \frac{N}{N_{c_i}}$$
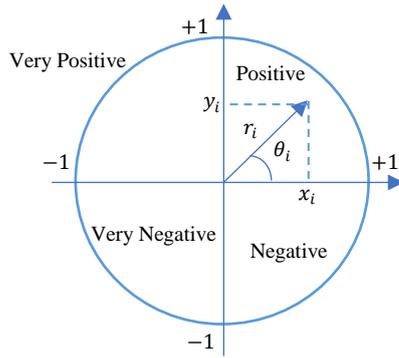(4)

$$r_i = TDOC(m, c_i)$$
(5)

Figure. 1 SentiCircle model in the polar coordinate system

Where, m is term and c_i is the context term. $f(c_i, m)$ is the number of times c_i occurs with m in sentences. N is the total number of terms, and N_{c_i} is the total number of terms that occur with c_i. And r_i is the radius word to centre origin in polar coordinate system.

2. Calculating prior sentiment score using sentiment lexicon. This score is used to calculate the angle term in the polar coordinate system.

$$\theta_i = Prior\_Sentiment(c_i) \times \pi \qquad (6)$$

3. Calculating point x_i and y_i, in which x represents sentiment strength of term and y represents sentiment orientation of term.

$$x_i = r_i \, cos\theta_i \qquad (7)$$

$$y_i = r_i \, sin\theta_i \qquad (8)$$

SentiCircle model that represents polar coordinate system is illustrated in Fig. (1):

To extract sentiment from given point in SentiCircle, SentiMedian is used. It calculates the minimum geometric median from all points. This geometric median determines the sentiment polarity of the context term. It can be positive, negative, or neutral. The SentiMedian equation, can be seen in Eq. (9).

$$g = min_{g \in R^2} \sum_{i=1}^{n} \|p_i - g\|^2 \qquad (9)$$

3. Research method

First, dataset will be pre-processed to clean terms from noise or less relevant information using pre-processing method. The pre-processed dataset is the input in module aspect categorization. To calculate word similarity, aspect keywords are needed as the input in this module. Aspect keywords are the result of keyword extraction using TF-ICF from Wikipedia links in Table 2. In aspect categorization, the candidate of aspect terms and opinion terms is extracted using grammatical rules. The aspect keywords and the candidate of aspect terms are categorized.

The next module is grouping dataset based on aspect to prepare for sentiment analysis using SentiCircle. Evaluation is conducted for aspect categorization and sentiment analysis to measure and to evaluate the proposed method of this research.

3.1 Data collection

The dataset is obtained from TripAdvisor website. This dataset had been annotated manually into 4 aspect categories. This dataset can contain multiple aspects in a review. The representation dataset can be seen in Table 3.

The researcher split the reviews into several sentences. Sentences that belong to the same review contain the same ID Review. The dataset can also have multiple categories. For instance, it can be seen in Table 4.

Table 3. Representation dataset review

ID Review	ID Sentence	Review
1004293	1	We, there were four of us, arrived at noon - the place was empty - and the staff acted like we were imposing on them and they were very rude.
1004293	2	The food was lousy - too sweet or too salty and the portions tiny.

Table 4. Representation dataset that contains multiple categories

Review	Category	Polarity
Service was divine, oysters where a sensual as they come, and the price can't be beat!!!	Service, Food	positive, positive

3.2 Pre-processing

The process taken for this step consists of:

Start 1. Taking the text review as an input; 2. If the module is Aspect Categorization, the pre-processing in Table 2 is performed with the sequence: Converting into Lowercase → Spelling Correction; 3. If the module is Keyword Extraction, the pre-processing in table 2 is performed with the sequence: Converting into Lowercase → Spelling Correction → Remove Punctuation

→ Stop-words Removal → Tokenization →
Lemmatization;
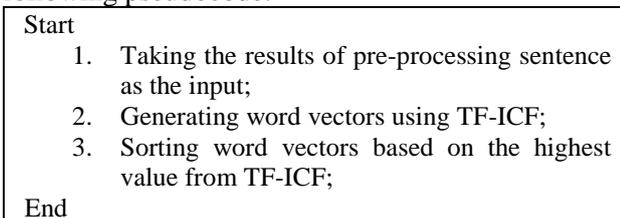4.   Save the results of pre-processing.
End
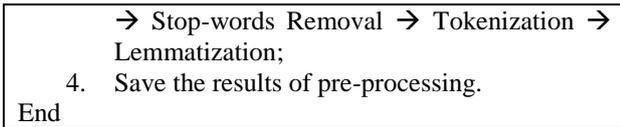
The pre-processing in module aspect categorization and keyword extraction is different. For the aspect categorization, the researcher used Stanford CoreNLP constituency parse for grammatical rule. Removing and changing any words from the sentence generated different result in constituency parse. In keyword extraction, stop-words removal, and lemmatization, it is needed to remove unimportant words, so the results only contain important keywords that represent the Wikipedia page.

3.3 Keyword extraction using TF-ICF

This module is used to generate aspect keywords using TF-ICF algorithm. From Fig. 2, there are two outputs from this module:
1.   Aspect keywords generated from Wikipedia;
2.   Aspect keywords generated from Wikipedia and external links.

The steps of this module explained in the following pseudocode:

Start
1.   Taking the results of pre-processing sentence as the input;
2.   Generating word vectors using TF-ICF;
3.   Sorting word vectors based on the highest value from TF-ICF;
End

3.4 Aspect categorization

The researcher conducted several experiments to determine the best aspect categorization (AC) performance. AC used grammatical rule to extract candidate aspect terms and opinion terms. The candidate aspect terms were calculated using word similarity with the aspect keywords to categorize them into the 4 pre-determined aspects in Table 1. The description of AC1, AC2 and AC3 can be seen as follows:
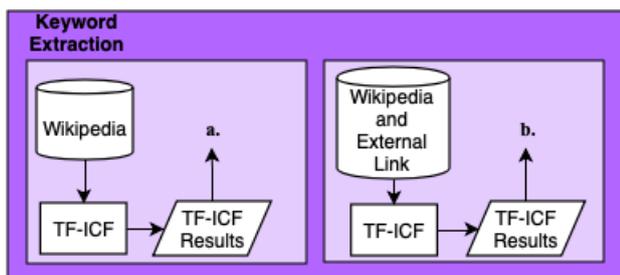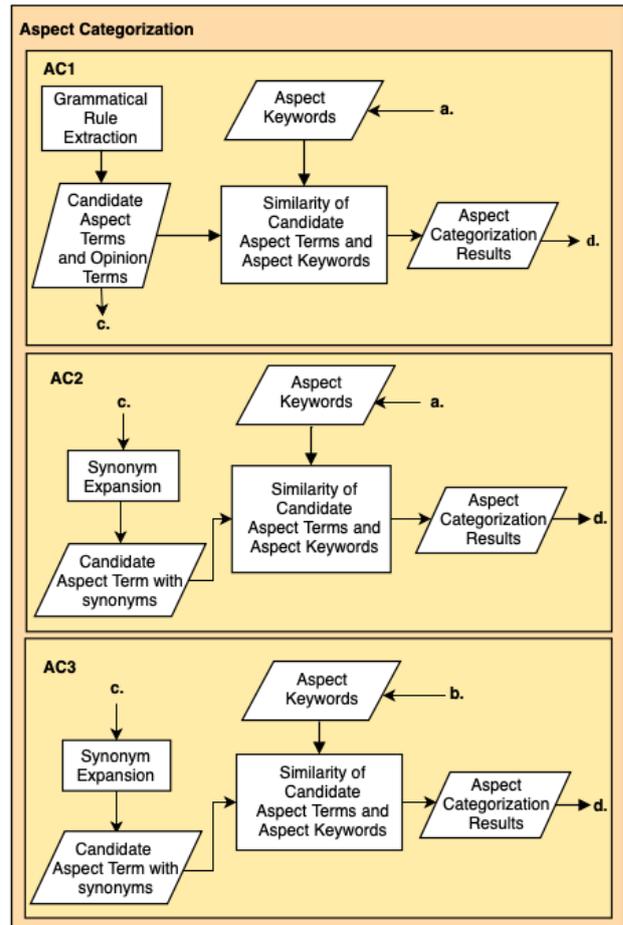

Figure. 2 Keyword Extraction Process

Figure. 3 Aspect Categorization Flow

3.4.1. Aspect categorization (AC) 1

This experiment is divided into the grammatical rule extraction and aspect categorization using word similarity. The description can be seen below:

Grammatical Rule Extraction

This process extracts the candidate aspect using Stanford CoreNLP [19]. Standford CoreNLP generates constituency parse tree that represents its syntactic structure from the sentence. To query the data from constituency parse tree, Tregex tool [20] is used. The step of this process explained in the following pseudocode:

Start
1.   Taking the results of pre-processing sentence as the input;
2.   Extracting clauses using Standford CoreNLP with Tregex pattern S < (NP $ VP);
3.   For each clause, Obtaining all phrases from constituency parse using Tregex pattern "NP", "ADVP", "ADJP", "VP", and "PP";
4.   For each phrase, the candidate aspects and the opinion words are extracted with the following steps:
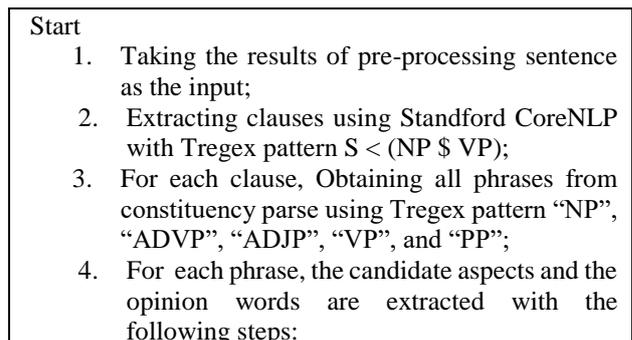
> a. Extracting the candidate aspects in NP;
> b. Extracting the opinion words in ADJP, ADVP, and VP;
> c. If the candidate aspects and opinion words are not found in the previous process, the candidate aspects and opinion words are found in PP.
> 5. For grammatically incorrect sentence, all nouns are extracted as the candidate aspects and all adjectives are extracted as the opinion words.
> End

The steps above generate the candidate aspects and the opinion words. The idea of this algorithm is to break down the sentence into several clauses, identify the candidate aspect in NP and PP, identify the opinions in NP, ADJP, ADVP, VP, and PP.

Aspect Categorization using Word Similarity

Word similarity is the last step in the aspect categorization process. The input of this process is the aspect keywords with the candidate aspect extracted using grammatical rules. This input is categorized into predetermined aspects in Table 1 using word similarity. In AC1, aspect keywords is generated from Wikipedia using TF-ICF.

To calculate the similarity, there is a possibility that the keywords overlap each category. To handle this, the researcher used TF-ICF value as the weighting to cosine similarity results. The formula of word similarity can be seen in Eq. (10), Eq. (11), and Eq. (12).

$$sem_{sim} = cosine_{sim} \times TF_{ICF} \qquad (10)$$

$$lex_{sim} = \begin{cases} 1, if\ a = k \\ 0, if\ a \neq k \end{cases} \qquad (11)$$

$$similarity = sem_{sim} + lex_{sim} \qquad (12)$$

From Eq. (11), a is the candidate aspect and k is the keyword. This equation aims to calculate the lexical similarity of words. From Eq. (12), the word similarity is the summation of semantic similarity and lexical similarity. The highest word similarity value was used as the class label of the document.

Table 5 shows that the yellow highlighted is the prediction of the aspect. For term list 1, the prediction aspect is service and for term list 2 the prediction aspect is food.

Table 5. Word similarity analysis results

Term List 1	Aspect 1 Food	Aspect 2 Service	Aspect 3 Ambience	Aspect 4 Price
Candidate Aspects: ['Service']. Prediction Aspect: Service				
Service	0.555	1.643	0.517	0.585
Total	0.555	**1.643**	0.517	0.585
Term List 2	Aspect 1 Food	Aspect 2 Service	Aspect 3 Ambience	Aspect 4 Price
Candidate Aspects: ['Vegetarian', 'Dish']. Prediction Aspect: Food				
Vegetarian	0.578	0.347	0.431	0.399
Dish	0.584	0.504	0.455	0.459
Total	**1.162**	0.851	0.886	0.858

Table 6. Word similarity expanded with synonyms

Term List 1	Aspect 1 Food	Aspect 2 Service	Aspect 3 Ambience	Aspect 4 Price
Candidate Aspects: ['Maintenance']. Synonyms: ['assistance', 'duty', 'utility'] Prediction Aspect: Service				
Maintenance	0.523	0.548	0.549	0.569
Assistance	0.419	0.681	0.406	0.465
Duty	0.421	0.485	0.475	0.462
Utility	0.479	0.516	0.460	0.543
Max	0.555	**0.681**	0.549	0.585

3.4.2. Aspect categorization (AC) 2

Aspect Categorization (AC) 2 undergoes the same process as AC1. However, the candidate aspect is expanded with their synonyms. The highest value is used as word similarity value.

From Table 6, the yellow highlighted is the highest value of word similarity between term list and aspect keyword, that is "assistance" meaning that the synonym words have the possibility value higher than the candidate aspect. This value is used to determine the aspect label from Table 1; and the aspect is Service.

3.4.3. Aspect categorization (AC) 3

Aspect Categorization (AC) 3 undergoes the same process as AC2. However, the aspect keywords used to perform the word similarity in the AC3 did not only use the term list from Wikipedia but also external links.

3.4.4. Baseline comparison

To evaluate our proposed method, the researcher compared this with the existing method for aspect extraction. The existing methods are:

1. Double Propagation (DP): DP is rule-based aspect extraction using dependency relation [21]. DP has pair of rules to extract word in direct dependency or indirect dependency. Direct dependency means that one word depends on other word without any additional words in their dependency relation. Indirect dependency means that one word depends on other word through some additional words. For dependency relation, there are 8 dependency relations are used to compose the rules, namely: amod, prep, nsubj, csubj, xsubj, dobj, iobj, and conj. In this research, comparison was done in the aspect extraction using DP rules. After the candidate aspect extracted, similarity was done using the same settings with the proposed method. To determine the best result, evaluation was carried out on the result of aspect categorization using DP + similarity with the proposed method;

2. DP[+]: DP with additional dependency relation to compose rules. DP[+] used 18 dependency relations [22]. This relations are: amod, prep, nsubj, csubj, xsubj, dobj, iobj, advmod, dep, cop, mark, nsubjpass, pobj, acomp, xcomp, csubjpass, poss, and conj. For comparison, it will use the same method with DP;

3. Latent Dirichlet Allocation (LDA): In LDA, the extracted candidate aspect is assumed as hidden topic [6]. In this research, gensim tool was used [23] to extract 600 hidden topics. If the reviews contain the word that exist in hidden topics then it is the candidate aspect. For similarity, it was used the same settings with the proposed method.

3.5 Pre-processed for aspect based sentiment analysis

At this stage, the data from the Aspect Categorization were categorized based on the same aspect categories. There are 4 datasets for each aspect. This dataset is the input of the next module.

3.6 Sentiment analysis

Several experiments were conducted to determine the best Sentiment Analysis (SA) performance. The experiments flow can be seen in Fig. 5. The description of the Sentiment Analysis experiment can be seen as follows.

3.6.1. Sentiment analysis (SA) 1

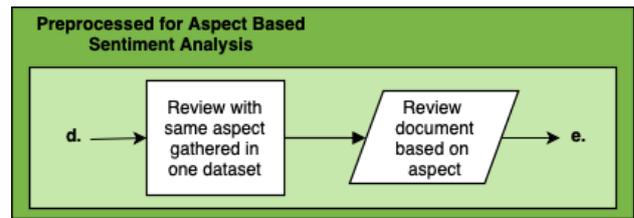The step of this experiment is explained in the following pseudocode:

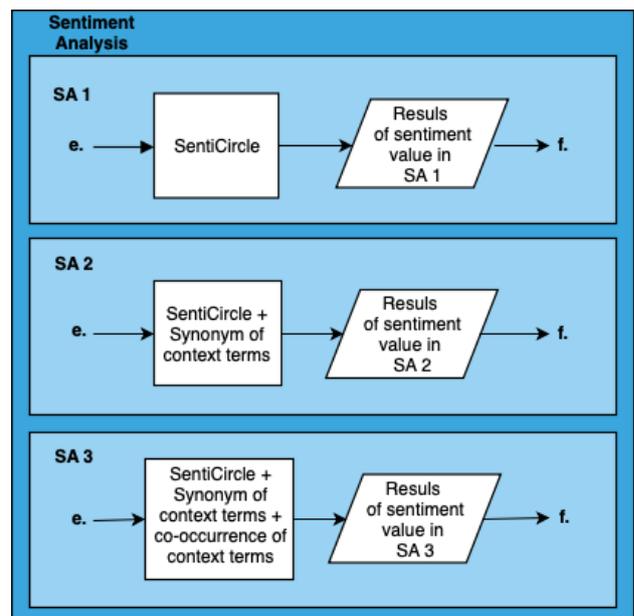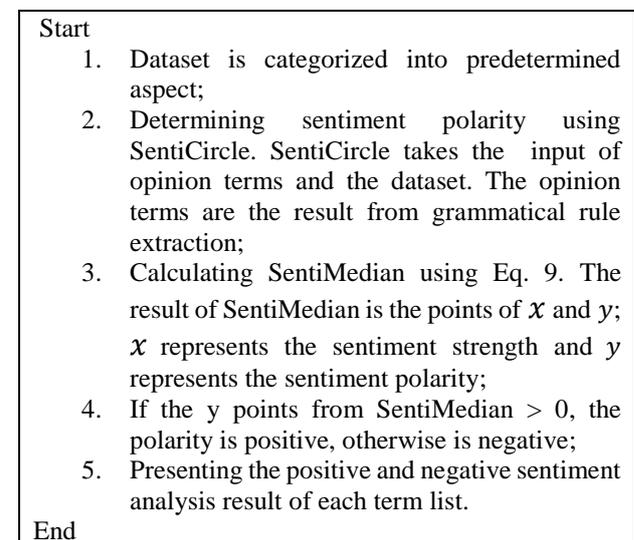Figure. 4 Pre-processed for aspect based sentiment analysis

Figure. 5 Flow of sentiment analysis

Start
1. Dataset is categorized into predetermined aspect;
2. Determining sentiment polarity using SentiCircle. SentiCircle takes the input of opinion terms and the dataset. The opinion terms are the result from grammatical rule extraction;
3. Calculating SentiMedian using Eq. 9. The result of SentiMedian is the points of x and y; x represents the sentiment strength and y represents the sentiment polarity;
4. If the y points from SentiMedian > 0, the polarity is positive, otherwise is negative;
5. Presenting the positive and negative sentiment analysis result of each term list.
End

The algorithm takes the opinion terms and dataset as the input for SentiCircle. The dataset is needed to compute the correlation of opinion terms with the context in the sentence. The sample of input data can be seen in Table 7, and the result can be seen in Table 8.

Table 7. Data input of SentiCircle

ReviewID	Review	Opinion Terms
1	I am not necessarily fanatical about this place, but it was a fun time for low prices.	[[fanatical], [fun, low]]
2	last but not least , prices for nyc are ridiculously low and the personnel is warm and friendly	[['ridiculously', 'low'], ['warm', 'friendly']]

Table 8. SentiCircle Result

Opinion Terms Review 1	SentiMedian		Prediction
	X	Y	
[fanatical]	0.461	-0.191	Negative
[fun, low]	0.125	0.251	Positive
Opinion Terms Review 2	SentiMedian		Prediction
	X	Y	
[ridiculously, low]	0.094	0.406	Positive
[warm, friendly]	0.177	0.456	Positive

Table 9. Context Words Expansion with Synonym

Opinion Terms	Context Terms	Synonym of Context Terms
[fanatical]	[low, necessarily]	[small, inferior, minor, naturally, automatically]
[fun, low]	[fanatical, low, necessarily, friendly, last, warm, ridiculously]	[small, inferior, automatically, naturally, passionate, enthusiastic, loving, welcoming, end, ending, sunny, hot, laughable]

3.6.2. Sentiment analysis (SA) 2

The process in SA2 is similar to SA1. However, in SA2 there is context terms expansion using synonym. The illustration of this expansion can be seen in Table 9.

3.6.3. Sentiment analysis (SA) 3

SA3 undergoes the same process with SA2 with additional expansion using co-occurrence context of terms. For instance in Table 9, the context terms of fanatical is low and necessarily. This context terms is expanded using co-occurrence in the dataset. The results of expansion are friendly, last, warm, and ridiculously.

3.6.4. Baseline comparison

To evaluate our proposed method for sentiment analysis. The result is compared with the existing lexicons:

1. SentiWordnet using first sense;
2. SentiWordnet using adapted lesk algorithm for word sense disambiguation [24].

3.7 Evaluation

Evaluation was conducted by comparing the number of performances of aspect categorization and sentiment analysis. Each performance was evaluated using *Precision*, *Recall*, and *F1-measure*. The formula of this methods is as follows.

$$Precision = \frac{TP}{TP+FP} \qquad (13)$$

$$Recall = \frac{TP}{TP+FN} \qquad (14)$$

$$F1\text{-}measure = 2 \times \frac{precision \times recall}{precision+recall} \qquad (15)$$

4. Result and analysis

In this section, the result of the Aspect and Sentiment analysis is explained.

4.1 Approach of aspect categorization

The evaluation results of aspect categorization are as follows.

Table 10 shows that AC3 is the best method. Thus, adding the external links can complement the existing Wikipedia keywords and can provide more precise predictions. For instance, review in AC1 is "The atmosphere is relaxed and casual"; the word atmosphere is predicted by system into Food category, while the ground truth of this word is Ambience. It is because the cosine similarity is greater towards Food, and the word has the same TF-ICF value with the keywords. In AC3, the system can predict correctly because there is a new keyword that does not exist in AC1 which has greater word similarity value. This keyword is "spirit". For more detail, It can be seen in Table 11.

Table 10. Result of aspect categorization

ASPECT CATEGORIZATION PERFORMANCE		
No. Approach	Performance Metrics	Average F1-Measure
AC1	Grammatical Rule + TF-ICF (Wikipedia) + Similarity	0.77
AC2	Grammatical Rule + Synonym + TF-ICF (Wikipedia) + Similarity	0.76
AC3	Grammatical Rule + Synonym + TF-ICF (Wikipedia and External Link) + Similarity	0.84

Table 11. Comparison of AC1 and AC3 with keyword "atmosphere"

Keyword AC1	Cosine Sim	TF-ICF	Semantic Sim	Prediction
food	0.4	1.0	0.4	Food
theme	0.37	1.0	0.37	Ambience
Keyword AC3	Cosine Sim	TF-ICF	Semantic Sim	Prediction
food	0.4	1.0	0.4	Food
spirit	0.51	0.98	0.5	Ambience
theme	0.37	1	0.37	Ambience

Table 12. Comparison of AC3 in each category

AC3 Performance on Each Aspect Category			
Aspect Category	Precision	Recall	F1-Measure
Ambience	0.86	0.80	0.83
Food	0.89	0.82	0.85
Price	0.86	0.73	0.79
Service	0.88	0.87	0.88

Table 13. Comparison of AC3 with the baseline

AC3 Performance with Baseline			
Method	Precision	Recall	F1-Measure
AC3	0.80	0.86	0.84
DP + Synonym + TF-ICF (Wikipedia and External Link) + Similarity	0.73	0.86	0.79
DP$^+$ + Synonym + TF-ICF (Wikipedia and External Link) + Similarity	0.71	0.89	0.79
LDA + Synonym + TF-ICF (Wikipedia and External Link) + Similarity	0.63	0.86	0.70

Next, the researcher conducted an analysis of the categories that had the best predictive value. This prediction was done on AC3, because AC3 was the best result of the system produced.

Table 12 shows that the Service category has the best performance compared to the others, while the worst is the Price. This is due to the fact that the similarity value from the keywords is less representative. For instance in the review, "This is the best sushi in new york city – hands down". Using grammatical rules, the candidate aspects "sushi", "city", and "hand" were extracted. From these candidate aspects, the system had predicted towards Price category. The system incorrectly predicted the word sushi, while the word ground truth is Food. Thus, the precision of the Price is low. For recall, it is because the grammatical rule cannot extract the candidate aspects in parallel clauses; for instance, in

review "The food is great and reasonably priced". From this review, the rule extracted the candidate aspect "food". The rule could not extract the candidate aspect "price" from that review. Thus, the recall is low because it did not predict correctly for Price category.

To find out the performance of the proposed method, the researcher also compared the result with the existing baseline method.

Table 13 shows that the proposed method is better than the existing baseline method which has the F1-Measure of 0.84. DP and DP$^+$ depend on dependency relation between words. They are working well in simple sentences, but in sentences that contain multiple clauses both of the methods cannot give accurate prediction; for instance in review, "I tend to judge a sushi restaurant by its sea urchin, which was heavenly at sushi rose". This review has ground truth Food category because of word "sushi". DP and DP$^+$ cannot extract the word "sushi" and the opinion word "heavenly" because the opinion word is located in different clauses with the candidate aspects.

DP+ uses more dependency relations than DP. This relations bring more correct aspects (higher recall) but also more errors (low precision). DP$^+$ also has higher recall than AC3, because of this addition relations; for instance in review "I loved everything about it-especially the shows and actors". AC3 cannot extract the candidate aspect words "shows" and "actors", because the nearest noun phrase from opinion word "loved" is "everything", but DP$^+$ can extract them through dependency relation advmod. One of the problem of AC3 that explained earlier is parallel clauses. DP$^+$ can extract them because of this addition relations.

For LDA, hidden topic doesn't represent the correlation between the candidate aspect and opinion word, because it generates based on word frequency; for instance in review "I have been coming here for years and have nothing but good things to say about the service and the great staff at La Lanterna". Using LDA, this review has hidden topics "come", "year", "service", and "staff". Using these hidden topics, the prediction of aspect category are Ambience and Service, while the ground truth is Service. Thus, it makes F1-Measure of LDA lower than AC3.

4.2 Approach of sentiment analysis

The evaluation results of sentiment analysis are as follows.

Table 14 shows that by expanding the context of words with synonyms and co-occurrence, the result of F1-measure is better. For instance, in the following

Table 14. Results of sentiment analysis

SENTIMENT ANALYSIS PERFORMANCE		
No. Approach	Performance Metrics	Average F1-Measure
SA1	SentiCircle	0.84
SA2	SentiCircle + Synonyms of context terms	0.84
SA3	SentiCircle + Synonyms of context terms + co-occurrence of context terms	0.87

Table 15. Comparison of SA3 with the baseline

SA3 Performance with Baseline			
Method	Precision	Recall	F1-Measure
SA3	0.83	0.92	0.87
SentiWordnet	0.87	0.62	0.73
SentiWordnet + WSD	0.85	0.68	0.75

reviews in the dataset:

1. The food is flavorful , plentiful and reasonably priced.
2. The shrimp scampi was excellent and the antipasti were plentiful.

The word "plentiful" from the reviews above was predicted into negative sentiment using SA1, while in SA3 predicted into positive sentiment. This is because the context words of "plentiful" in SA1 is "flavorful" and "excellent". In SA3, the context words of "flavorful" and "excellence" are extended using synonyms and co-occurrence. Thus, SA3 can produce more accurate predictions. It can be seen in Fig. 6
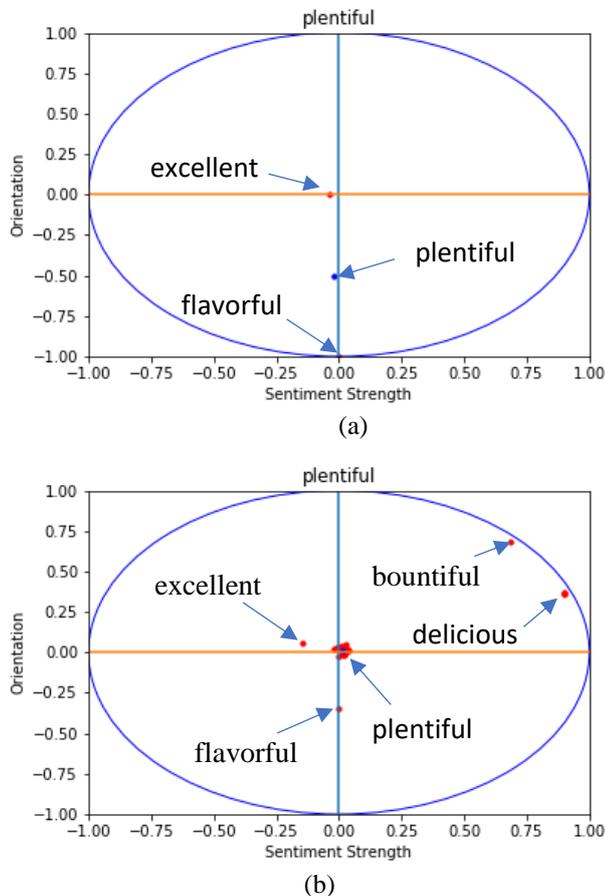
Table 16. Sentiment evaluation results on restaurant

Sentiment Evaluation Results on Aspect		
Aspect	Sentiment	Evaluation Results (in Percent)
Ambience	Positive	24.62
	Negative	1.215
Food	Positive	37.89
	Negative	3.748
Service	Positive	16.01
	Negative	1.621
Price	Positive	8.814
	Negative	6.079
Total Percentage		100

To find out the performance of the proposed method in Sentiment Analysis, the researcher compared the result with the existing baseline method. The results can be seen in Table 15.

Table 15 shows that the proposed method is better than the existing baseline method which has the F1-measure of 0.87.

4.3 The results of aspect based sentiment analysis

The evaluation results of aspect based sentiment analysis for restaurant review is as follows.

Table 16 shows that the restaurant has positive reviews in each aspect.

4.4 Effect of aspect to the sentiment

SentiCircle polarity can be obtained based on the context. For instance, there are several reviews for Food category as follows.

1. The takeout is great too since they give high quality tupperware as well;
2. I highly recommend Caviar Russe to anyone who wants delicious top grade caviar and fantastic service;
3. I highly recommend the Sophia pizza;
4. I have to highly recommend the lobster roll - not too much mayo; you can tell it was a fresh lobster.

The reviews for Price category are as follows

1. The high prices you're going to pay is for the view not for the food;

(a)

(b)

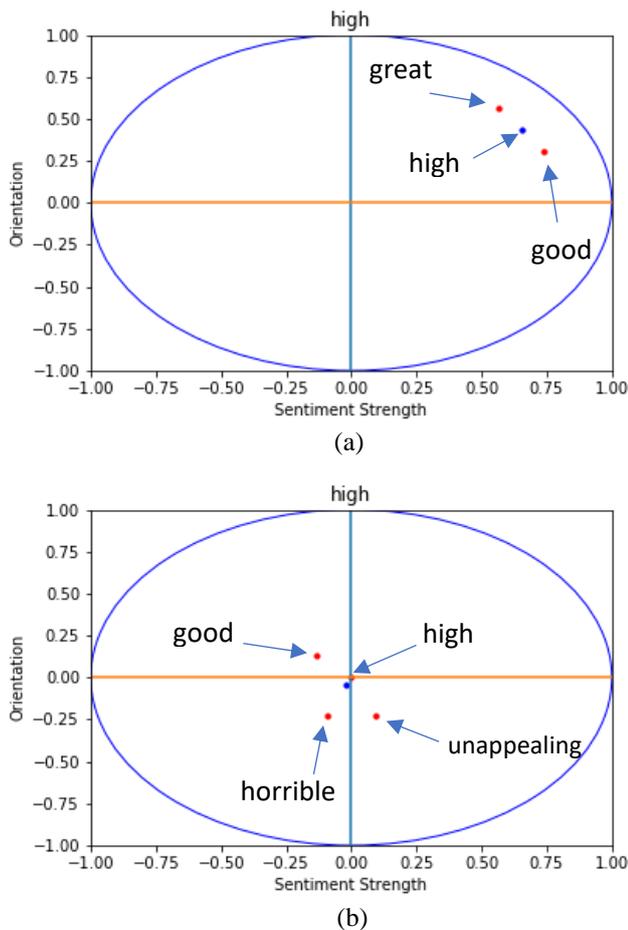Figure. 6 SentiMedian of word "plentiful": (a) SA1 and (b) SA3

(a)

(b)

Figure. 7 SentiMedian of word "high" in : (a) Food category and (b) Price category

2. Price is high but the food is good, so I would come back again;
3. Prices too high for this cramped and unappealing restaurant;
4. Even though its good seafood, the prices are too high.

The word "high" from the reviews in Food category has different polarity based on the context words. Fig. 7 shows that sentiment polarity is positive for Food category and negative for Price category.

5. Conclusion and future works

This research proposes aspect based sentiment analysis using grammatical rules, word similarity, and SentiCircle. Grammatical rules and word similarity are used for aspect categorization. SentiCircle is used for sentiment analysis. The best performance for aspect categorization is AC3 with F1-Measure of 0.84. AC3 uses keyword from Wikipedia and external links. For sentiment analysis, the best performance is SA3 with F1-Measure of 0.87.

SA3 expands the context word of SentiCircle using synonym and co-ocurrence word.

In this research, SentiCircle can change the polarity sentiment of a review based on the context. It can be seen from the experiment that the word "high" in the Food category has different sentiment polarity from the Price category. For future work, the method aspect extraction can be enhanced to solve the parallel clauses, exclamation opinion sentences, and sentence that contains implicit aspect or implicit opinions.

Acknowledgments

References

[1] X. Xu, X. Wang, Y. Li, and M. Haghighi, "Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors", *International Journal of Information Management*, Vol. 37, No. 6, pp. 673–683, 2017.

[2] B. Liu, "Sentiment Analysis and Opinion Mining", *Synthesis Lectures on Human Language Technologies*, Vol. 5, No. 1, pp. 1–167, 2012.

[3] Suhariyanto, A. Firmanto, and R. Sarno, "Prediction of Movie Sentiment based on Reviews and Score on Rotten Tomatoes using SentiWordnet", In: *Proc. of International Seminar on Application for Technology of Information and Communication*, pp. 202–206, 2018.

[4] B. S. Rintyarna, R. Sarno, and C. Fatichah, "Enhancing the performance of sentiment analysis task on product reviews by handling both local and global context", *International Journal of Information and Decision Science*, Vol. 11, 2018.

[5] D. Khotimah and R. Sarno, "Sentiment Analysis of Hotel Aspect Using Probabilistic Latent Semantic Analysis, Word Embedding and LSTM", *International Journal of Intelligent Engineering and Systems*, Vol. 12, No. 4, pp. 275–290, 2019.

[6] R. Priyantina and R. Sarno, "Sentiment Analysis of Hotel Reviews Using Latent Dirichlet Allocation, Semantic Similarity and LSTM",

International Journal of Intelligent Engineering and Systems, Vol. 12, No. 4, pp. 142–155, 2019.

[7] K. Schouten and F. Frasincar, "Survey on Aspect-Level Sentiment Analysis", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 28, No. 3, pp. 813–830, 2016.

[8] M. Hu and B. Liu, "Mining and summarizing customer reviews", In: *Proc. of International Conference on Knowledge Discovery and Data Mining*, p. 168, 2004.

[9] G. Carenini, R. T. Ng, and E. Zwart, "Extracting knowledge from evaluative text", In: *Proc. of International Conference on Knowledge Capture*, pp. 11–18, 2005.

[10] L. H. Suadaa and A. Purwarianti, "Combination of Latent Dirichlet Allocation (LDA) and Term Frequency-Inverse Cluster Frequency (TFxICF) in Indonesian text clustering with labeling", In: *Proc. of International Conference on Information and Communication Technology*, pp. 1–6, 2016.

[11] H. Saif, Y. He, M. Fernandez, and H. Alani, "Contextual semantics for sentiment analysis of Twitter", *Information Processing and Management*, Vol. 52, No. 1, pp. 5–19, 2016.

[12] P. D. Turney and P. Pantel, "From frequency to meaning: Vector space models of semantics", *Journal of Artificial Intelligence Research*, Vol. 37, No. 5, pp. 141–188, 2010.

[13] S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0 : An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining", In: *Proc. of International Language Resources and Evaluation*, pp. 1–12, 2010.

[14] F. Nurifan, R. Sarno, and C. S. Wahyuni, "Developing corpora using word2vec and wikipedia for word sense disambiguation", *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 12, No. 3, pp. 1239–1246, 2018.

[15] E. Faisal, F. Nurifan, and R. Sarno, "Word Sense Disambiguation in Bahasa Indonesia Using SVM", In: *Proc. of International Seminar on Application for Technology of Information and Communication*, pp. 239–243, 2018.

[16] B. S. Rintyarna and R. Sarno, "Adapted weighted graph for Word Sense Disambiguation", In: *Proc. of International Conference on Information and Communication Technology*, pp. 1–5, 2016.

[17] W. H.Gomaa and A. A. Fahmy, "A Survey of Text Similarity Approaches", *International Journal of Computer Applications*, Vol. 68, No. 13, pp. 13–18, 2013.

[18] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information", *Transactions of the Association for Computational Linguistics*, Vol. 5, pp. 135–146, 2017.

[19] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky, "The Stanford CoreNLP Natural Language Processing Toolkit", In: *Proc. of Association for Computational Linguistics: System Demonstrations*, pp. 55–60, 2014.

[20] R. Levy and G. Andrew, "Tregex and Tsurgeon: tools for querying and manipulating tree data structures", In: *Proc. of International Conference on Language Resources and Evaluation*, 2006.

[21] G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion Word Expansion and Target Extraction through Double Propagation", *Computational Linguistics*, Vol. 37, No. 1, pp. 9–27, 2011.

[22] Q. Liu, Z. Gao, B. Liu, and Y. Zhang, "Automated rule selection for opinion target extraction", *Knowledge-Based Systems*, Vol. 104, No. 15, pp. 74–88, 2016.

[23] R. Rehurek and P. Sojka, "Software Framework for Topic Modelling with Large Corpora", In: *Proc. of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pp. 45–50, 2010.

[24] S. Banerjee and T. Pedersen, "An Adapted Lesk Algorithm for Word Sense Disambiguation Using WordNet", In: *Proc. of International Conference on Intelligent Text Processing and Computational Linguistics*, pp. 136–145, 2002.