Received: August 5, 2019

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Aspect Based Sentiment Analysis for Restaurant Reviews Using Hybrid ELMo-Wikipedia and Hybrid Expanded Opinion Lexicon-SentiCircle

Farza Nurifan[1] **Riyanarto Sarno[1]*** **Kelly Rossa Sungkono[1]**

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

Abstract: Many restaurant review analysis have been done, however only few analysis have been done for specific aspects of a restaurant. In this context this paper proposes aspect based restaurant analysis which includes Physical environment, Food quality, Service quality and Price fairness. The analysis steps include Aspect Term Extraction (ATE), Aspect Keyword Extraction (AKE), Aspect Categorization (AC) and Sentiment Analysis (SA). ATE employs the modification of Double Propagation method and several Topic Modelling methods, AKE utilizes Term Frequency-Inverse Cluster Frequency (TF-ICF), in AC we propose Hybrid ELMo-Wikipedia (HEW), and in SA we propose Hybrid Expanded Opinion Lexicon-SentiCircle (HEOLS). The results show that our modification of the methods used in ATE could increase the f1measure of the AC by average 2%, then the HEW that we proposed had better f1measure compared to other similar methods by average 6%. Other than that, our proposed HEOLS can expand and redetermine the Opinion Lexicon polarity and can increase f1measure of SA by 6%.

Keywords: Aspect based sentiment analysis, Natural language processing, Opinion mining, Sentiment analysis, ELMo, SentiCircle.

1. Introduction

The development of digital technology allows people to express their opinions about a restaurant through the internet by writing a review. Review written by a person is very useful for restaurant business owners to evaluate the quality of their restaurant. One way to represent a review is with Aspect based Sentiment Analysis (ABSA). Unlike sentiment analysis, ABSA can represent reviews in more detail.

ABSA's previous research related to this paper was carried out by García-Pablos et al. [1], in restaurant reviews they used three aspects, namely Food, Service and Ambiance. In this study we propose ABSA with aspects in the form of four criteria that determine restaurant quality from Gagić et al. [2] namely Physical Environment, Food quality, Service quality and Price fairness.

In this study we conducted ABSA for restaurant reviews in four stages, namely Aspect Keyword Extraction (AKE), Aspect Term Extraction (ATE), Aspect Categorization (AC) and Sentiment Analysis (SA). ATE is the process of taking a word in a review that indicates an aspect, the word is called aspect term. AKE is the process of taking words that represent an aspect, which is called aspect keywords. AC is the process of determining a review into which category based on aspect term and aspect keyword extracted in ATE and AKE. SA is a process to determine the sentiments of aspects that are in a review.

Previous research on ATE was carried out using the Double Propagation (DP) [3] and Topic Modeling [4, 5] methods. However, this method has drawbacks where there are still reviews that the aspect terms are not extracted. As an example, the aspect terms that are not extracted using DP is in the sentence "The sushi seemed pretty fresh and was adequately proportioned", the DP algorithm does not produce any output even though there is aspect term "sushi" in the sentence. Similar to DP, the methods of Topic modeling used in [4, 5] also have many sentences that the aspect terms are not extracted. Other than that, the extracted aspect terms are not always correct. For example, in sentence "I went to this place on Saturday and it was amazing" the extracted aspect term should only be "place" but the word "Saturday" which is the name of the day (not an aspect term) is also taken.

ABSA conducted by Akhtar et al. [6] at the AKE stage they determine the aspect keywords one by one manually. That is certainly difficult to do and also requires considerable effort.

In AC, the previous study [4, 5] used a semantic similarity calculation between vector keyword with vector aspect term. They got the vector value from word embedding Glove. The vector values that exist in Glove are static, meaning that from their semantic similarity calculations, an aspect term will always fit into the same aspect. For example, in the sentences "This food is delicious" and "This food is expensive", both have aspect term "food" but have different aspects depending on the context. The first sentence discusses the taste of food (Food quality) while the second sentence discusses about the food prices (Price fairness).

For SA, in previous research by Hu et al. [7] they use Opinion Lexicon. Opinion Lexicon is a bag-of-words of positive opinion words and negative opinion words. Because it is bag-of-word, the polarity value of an opinion word becomes static even though one opinion word can have different polarity depending on the aspect. For example, the word "cheap" on Opinion Lexicon the value is negative but if the word is in the aspect of Price fairness then it should be positive. Other than that, the weakness of the Opinion Lexicon is that not all opinion words are in bag-of-words of positive words or negative words, for example the word "flavorful" in the sentence "the food is flavorful" should be positive, or the word "high" the sentence "the prices are too high" should be negative, but both of them are not in Opinion Lexicon. Because the words "flavorful" and "high" are not in the Opinion Lexicon, they have no polarity value and that makes the results of SA always negative.

Other related research is done by Firmanto et al. [8]. They focus on developing methods for extracting aspects terms (ATE stage). In AC, they only use existing methods similar from previous research in [4] and [5] without modification or addition, the difference is they used Fasttext instead of Glove. However, Fasttext and Glove have the same characteristics, so they still have lacks that we have explained in paragraph 6. Also in SA they use SentiCircle [9] with a few additions but still have lacks as we explained in the paragraph 7.

Of all the shortcomings at each stage of the previous study we proposed several methods to improve it. For ATE, we modified the method in previous research, namely Double Propagation and Topic Modeling with Noun Extraction and Aspect Term Filter based on Part-of-Speech (POS) Tag and Named Entity Recognition so that it can handle reviews that the aspect terms are not extracted and handles extracted aspect term errors. At AKE we propose a semi-supervised method for extracting aspect keywords using Term Frequency-Inverse Cluster Frequency (TF-ICF) and data from Wikipedia that can reduce the effort of extracting aspect keywords manually one by one. Then for AC, we propose Hybrid ELMo-Wikipedia which can determine vectors for aspect terms and aspect keywords based on the context while Firmanto et al. [8] that only use existing proposed method from Priyantina et al. [4] and Khotimah et al. [5]. Finally, for SA, rather than using SentiCircle with some additions like Firmanto et al. [8] we propose Hybrid Expanded Opinion Lexicon-SentiCircle which can change the polarity of words in the Opinion Lexicon based on aspects and can expand Opinion Lexicon so that there are no words that do not have polarity values.

Furthermore, section 2 contains the theories related to this research, section 3 contains the methods we propose, section 4 contains the results and analysis of the experiments we conducted, and finally section 5 contains conclusions.

2. Related theory

This section explains the theories related to this research.

2.1 Restaurant aspect

Based on a literature study by Gagić et al [2] from previous researchers who discussed about customer expectation and service-quality perception in the food service industry, the quality of restaurants is categorized into four dimensions namely Food quality, Service quality, Physical environment, and Price fairness. We use these four dimensions as aspects. The variable that determine the aspects can be seen in Table 1.

2.2 Double propagation

Double Propagation (DP) is an algorithm proposed by Qiu et al. [3] to extract aspect terms using the rules they create. These rules are based on dependency relation. Dependency relation is grammatical relationship between words in a sentence [10]. To extract dependency relation from a sentence, they use the MiniPar tool. In this study we use the CoreNLP tool to parse the data, so we need to convert POS tags and dependency relations from MiniPar to CoreNLP. To convert this, we followed what was done in the previous research in [11].

2.3 Topic modelling

In Natural Language Processing (NLP) Topic Modeling is a very useful method for finding topics and finding semantic relations among many unstructured documents [12]. There are many Topic Modeling methods used by researchers, including Latent Dirichlet Allocation (LDA) [13], Probabilistic Latent Semantic Analysis (PLSA) [14], and LDA2Vec [15].

Table 1. Aspect variables

Aspect	**Variable**
Physical environment	ambience, table settings, facility aesthetics, decor, lighting, layout, and service staff appearance.
Food quality	menu variety, healthy options, nutrition, food is served at the appropriate temperature, serving size, food presentation is attractive, menu design, tastiness of food, freshness
Service quality	the chain restaurant brand has my best interests at heart, employees are always willing to help me, attentive stuff, staff appearance, employees have the knowledge to answer my questions, friendly dining managers
Price Fairness	overall value of the dining experience, reasonable price items, good value for money

Table 2. Preprocessing description

Preprocessing	**Description**
Lowercase	This process converts all capital letters that are in a text into lowercase letters.
Remove punctuation	This process removes all punctuation in the text.
Remove stopwords	This process removes all stopwords that are in a text.
Lemmatization	This process changes the words in a text into its basic form (lemma).
Minimum word limit	This process removes all words in a text consisting of three characters or less.

2.4 Preprocessing

Preprocessing is a process to eliminate disturbances that are in the text [16]. There are several preprocessing conducted in this study that can be seen in Table 2.

2.5 TF-ICF

Term Frequency Inverse Cluster Frequency (TF-ICF) [17] is a term/word weighting method based on the frequency of occurrence in many document clusters. TF-ICF can be used to search for important words that are in many document clusters. The equation of TF-ICF can be seen in Eqs. (1) and (2).

$$ICF_x = \log\left(\frac{N}{CF_x}\right) \tag{1}$$

$$TF - ICF_{x,i} = TF_{x,i} \times ICF_x \tag{2}$$

where,
N = Total clusters
CF_x = Total clusters that contains term x
$TF_{x,i}$ = Total term x in cluster i

2.6 ELMo

ELMo is deep contextualized word representation [18]. In contrast to word representations such as Glove [19], Fasttext [20], and Word2vec [21] where a word has only one vector representation, with ELMo a word can have many vector representations depending on the context. For example, the word “bucket” in the following three sentences:

“He dropped the bucket.”
“I have a bucket list to do.”
“The bucket was filled with oil.”

The word "bucket" in the second sentence has different meanings with the words in the first and third sentences. But it will be considered the same word using Glove, Fasttext, and Word2vec. With ELMo the bucket word is represented by different vectors because the words (context) that are around the bucket word are different.

In this study we used TensorFlow tools [22] to implement ELMo. Input from ELMo can be in the form of words, arrays of words, sentences, or arrays of sentences. If the input is word, the output is a 1024-dimensional vector. If the input is in the form of a sentence, the output is an array of 1024-dimensional vector as many words as there are in the sentences entered.

2.7 Semantic similarity

Semantic similarity is a distance calculation that is not calculated lexically but based on the meaning of the word [23]. There are many ways to calculate semantic similarity, one of them using cosine similarity. Cosine similarity works by measuring the angle between two vectors [24]. The cosine similarity equation can be seen in Eq. (3).

$$cosine(w_a, w_b) = \frac{\sum_{i=1}^{n} w_{ai} w_{bi}}{\sqrt{\sum_{i=1}^{n} w_{ai}^2}\sqrt{\sum_{i=1}^{n} w_{bi}^2}} \quad (3)$$

where,
w_{ai} = vector member from w_a
w_{bi} = vector member from w_b

2.8 Named entity recognition

Named Entity Recognition (NER) is a method for finding entity names such as people's names, organization names, locations, times, etc. [25]. One of the tools that can be used easily is NER at Stanford CoreNLP [26]. There are three types of NER in Stanford CoreNLP, namely:

1. Name (Person, Location, Organization, Misc)
2. Number (Money, Number, Ordinal, Percent)
3. Time (Date, Time, Duration, Set)

2.9 Opinion Lexicon

Opinion Lexicon is a list of opinion words containing a list of positive opinion words (positive opinion lexicon) and a list of negative opinion words (negative opinion lexicon). The Opinion Lexicon used in this study is the Opinion Lexicon used by Qiu et al. [3]. The Opinion Lexicon originally comes from Hu et al. [7] which continues to grow to date with around 6800 words.

2.10 SentiCircle

SentiCircle is a method for determining sentiment based on context [9]. They represent sentiment in the form of polar coordinates (Fig. 1) where the y axis represents the value of the sentiment polarity and the x axis represents the strength of sentiment. To determine the polarity value of a word first calculate the Term Degree of Correlation (TDOC) which is the degree of correlation between the term m and the c_i context in a document using Eq. (4). Next determine the radius for each $c_i$context with Eq. (5). Then determine the angle of each c_i context uses Eq. (6). Finally determine the x and y positions for each c_i context using Eqs. (7) and (8).

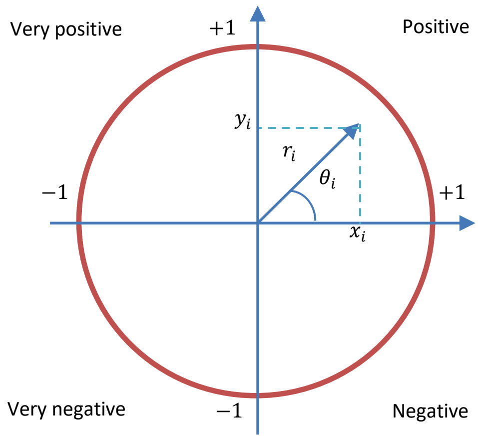

Figure. 1 SentiCircle representation

The final result of the polarity value of a word is determined by calculating the median of all positions in the existing word c_i context. They [9] named the determination of polarity using the median with the name SentiMedian. SentiMedian calculation is done by Eq. (9).

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

$$r_i = TDOC(m, c_i)_d \quad (5)$$

$$\theta_i = Prior\_Sentimen(c_i) \times \pi \quad (6)$$

$$x_i = r_i cos\ \theta_i \quad (7)$$

$$y_i = r_i sin\ \theta_i \quad (8)$$

$$g = \arg\min_{g \in R^2} \sum_{i=1}^{n} ||p_i - g||_2 \quad (9)$$

where,
d = Document
m = Opinion word
c_i = Opinion word context
N = Number of m in d
N_{c_i} = Number of c_i in d
$f(c_i, m)$ = The frequency of joint occurrences between m and c_i in d
r_i = Radius
$Prior\_Sentimen$ = polarity value
θ_i = c_i degree (in radian)
x_i and y_i = c_i position in x axis and y axis
p_i = c_i position
g = Position of m (x_m and y_m)

3. Research method

The stages of the method we propose can be seen in Fig. 2. In the picture there are four general stages, namely Aspect Term Extraction, Aspect Keyword Extraction, Aspect Categorization and Sentiment Analysis.

3.1 Dataset

The dataset we used in this study was from SemEval 2016 Task 5 [27] Subtask 1 about restaurant reviews. The dataset contains reviews with many sentences. The sentences in the dataset are labeled manually by professional annotators. From the labels in the dataset we then categorize it into four aspects categories according to the one in Table 1. The dataset labels that are in accordance with these aspects are seen in Table 3.

In this study we did not handle review sentences with implicit aspects (review sentences that did not have at least one aspect term). In the dataset we use it can be in one sentence review has many aspects. For data with two or more sentiment labels in the same aspect category, sentiment labels are determined based on the highest number of sentiment labels. If the amount is the same, it will be considered neutral. Examples of the dataset we use can be seen in Table 4.

The distribution of aspect categories in the dataset we use is not balanced. Therefore, we balance the data using the undersampling technique. We did the undersampling by deleting the data in the majority categories (Food quality, Physical environment, and Service quality) as can be seen in Table 5.

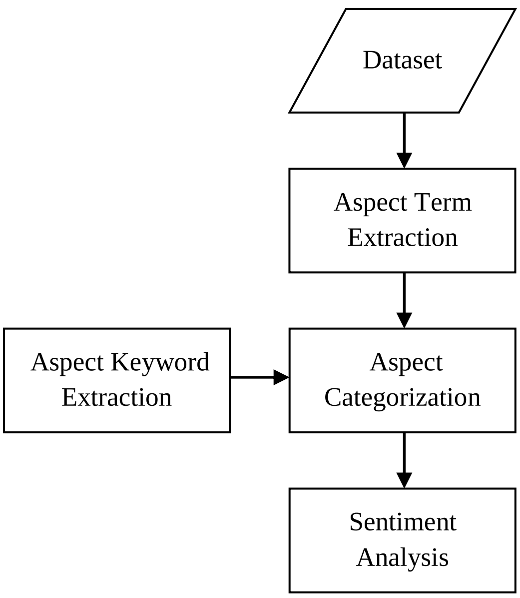

Figure. 2 Method stages

Table 3. Coresponding dataset label to aspect

Aspect	Corresponding dataset label
Physical environment	ambience#general location#general
Food quality	food#quality food#style_options drinks#quality drinks#style_options
Service quality	service#general
Price Fairness	drinks#prices restaurant#prices food#prices

Table 4. Dataset examples

Review	Aspect	Sentiment
Great pizza and fantastic service.	Food quality, Service quality	Positive, Positive
There was a small wait, but shorter than I expected.	Service quality	Positive
Located at the end of a magnificent block.	Physical environment	Positive
Drinks way over priced.	Price Fairness	negative

Table 5. Data distribution

Aspect	Data distribution before balancing	Data distribution after balancing
Physical environment	16%	23%
Food quality	51%	33%
Service quality	23%	24%
Price Fairness	10%	20%

The distribution of aspect categories in the dataset we use is not balanced. Therefore, we balance the data using the undersampling technique. We did the undersampling by deleting the data in the majority categories (Food quality, Physical environment, and Service quality) as can be seen in Table 5.

3.2 Aspect term extraction (ATE)

The existing method used for the first ATE stage is using Double Propagation then the second, third, and fourth method are using topic modeling LDA, PLSA, and LDA2Vec. For DP we do preprocess with lowercasing all the data. Then for Topic Modeling we try to replicate as closely as possible from the method proposed by Priyantina et. al. [4] and Khotimah et. al. [5] where they use LDA and PLSA respectively to extract the aspect terms. The difference between their methods and this research is the preprocessing data. In here we use the preprocessing technique we use in AKE. Then, we modify all the methods above with Noun Extraction and Aspect Term Filter.

3.2.1. Noun extraction

We overcome sentences that does not have aspect terms by taking all the noun words in the sentence as aspect terms. We take the word noun using TokensRegex [28]. TokensRegex is a tool to query the word attributes in a sentence such as lemma, tag (POS Tag), normalized, and so on. The TokensRegex query that we use to take the noun word is "[{post: NN}] | [{post: NNS}] | [{post: NNP}] | [{post: NNPS}]" without quotes.

3.2.2. Aspect term filter

To reduce the wrong aspect terms that are extracted we overcome it by filtering the extracted aspect terms using Named Entity Recognition (NER). Aspect terms that have a NER tag other than the NER tag “Title” will be deleted. We didn't delete NER tag “Title” because of the name of a title such as waiter, waitress, manager, DJ, and the other possibility is the correct term. For example, in the sentence "The waiter was very nice to me", in the sentence the waiter was indeed an aspect term. Finally, we filter the aspect terms that contain the string "time" because the filter from NER only removes the timepiece without the word "time" itself and the aspect term that contains the string "thing" because it is a general word for expressing things so it's clearly not an aspect term.

3.3 Aspect Keyword Extraction (AKE)

At this stage we extract the aspect keywords using TF-ICF with data from Wikipedia and some preprocessing techniques.

3.3.1. Wikipedia

In this study, we propose several Wikipedia pages that correspond to the four aspects we have explained in Table 1. The Wikipedia page that we selected can be seen in Table 6.

Table 6. Wikipedia pages

Aspect	**Title of Wikipedia pages**
Physical environment	Theme restaurant, Atmosphere (architecture and spatial design), Ambience (sound recording)
Food quality	Food, Drink, Meal
Service quality	Customer service, Waiting staff
Price Fairness	Price, Pricing

3.3.2. Preprocessing

Before data from Wikipedia is calculated using TF-ICF, we preprocess the data first. Preprocessing is done that is Lowercase, removing punctuation, removing stopwords, Lemmatization, and the minimum word limit.

3.3.3. TF-ICF

The next aspect of the keyword extraction process is done using TF-ICF. The algorithm can be seen in Fig. 3.

3.4 Aspect categorization (AC)

We categorize aspects using the calculation of semantic similarity between aspect keywords and aspect terms. The values of the aspect keywords and aspect terms are determined using ELMo and Wikipedia.

3.4.1. Hybrid ELMo-Wikipedia (HEW)

We propose Hybrid ELMo-Wikipedia to determine word vector values from aspect terms and keyword aspects. Wikipedia data used are the same as those used for keyword extraction which can be seen in Table 6.

a) Aspect keyword vector determination

The algorithm for determining the aspect keyword vector can be seen in Fig. 4. The average calculation in line 9 from Fig. 4 is done with Eqs. (10) and (11).

$$\bar{v} = \{v_1, v_2, \ldots, v_n\} \quad (10)$$

$$v_n = \frac{\sum_{i=1}^{k} v_{nk}}{k} \quad (11)$$

where,
$\bar{v}$ = Vector of aspect keyword
v_n = Vector member of $\bar{v}$
n = Vector dimension
k = Number of all vector words

Start
1. Retrieve Wikipedia data that has been preprocessed as input
2. Calculate the Term Frequency (TF)
3. Calculate the ICF value using Eq. (1)
4. Calculate the TF-ICF value using Eq. (2)
5. Take ten words with the highest TF-ICF value as the aspect keywords

End

Figure. 3 Keyword extraction using TF-ICF

```
Start
1.  For each aspect keyword ak of aspect A
2.     For each sentence s in Wikipedia pages
          according to A
3.        For each word w in s
4.            If ak in w
5.                Input s to ELMo
6.                Get w vector from ELMo
                  output
7.        Endif
8.     Endfor
9.     Calculate the average of all w vector
10.    Assign calculation results as the aspect
       keyword vector
11. Endfor
End
```

Figure. 4 Aspect keyword vector determination

b) Aspect term vector determination

To determine the vector aspect term extracted at the ATE stage, we use the review sentence from the aspect term as input for ELMo. Then we take the aspect term word vector from the ELMo as the aspect term vector.

3.4.2. Semantic similarity

In this part, semantic similarity is calculated to determine the category categories of aspect term taken at the Aspect Term Extraction stage. Calculation of similarity is done by Eq. (11).

$$Similarity(A,t) = \frac{\sum_{i=0}^{n} cosine(A_i, t)}{n} \tag{11}$$

where,

A = Aspect

A_i = Aspect keyword vector of A

t = Aspect term vector

n = Number of aspect keywords that are in A

$cosine()$ = Eq. (3)

Aspect categories are determined based on the highest value of the results of semantic similarity calculations.

3.4.3. Special case

In this part we add the Named Entity Recognition (NER) feature to the Aspect Categorization. We immediately categorize sentences containing words with the NER tag "Money" into the category Price fairness. We also include that word with NER tag "Money" as an aspect term. We do that because everything that talks about money is almost certainly in the Price fairness category.

3.4.4. Comparison

We compare the Hybrid ELMo and Wikipedia methods that we propose for the determination of aspect term vector and aspect keyword vector with the method of determining the aspect term vector and aspect keyword vector from previous research using Glove [4, 5, 29], Word2vec [29] and Fasttext [8]. In addition, we also compare with ELMo without modification.

3.5 Sentiment analysis

At this stage the review data that has been extracted aspect terms and categorized aspects will be determined sentiment. In this study we only use data labeled positive or negative sentiments, we do not use data with neutral sentiment labels because the amount of the data is too little.

3.5.1. Opinion term extraction

We determine the opinion word from an aspect by taking the adjective word closest to the aspect term. The distance between adjective words and aspect terms is determined based on the number of words in between. For example, in the phrase "Very good place and the pizza is amazing", in the sentence there are two aspects, namely "place" and "pizza". The opinion word for the term "place" is "good" because the word is the adjective word that is closest to the aspect term "place". Then the same with the aspect of the term "pizza" the opinion word is "amazing".

3.5.2. Hybrid expanded opinion Lexicon-SentiCircle (HEOLS)

We propose the Hybrid Expanded Opinion Lexicon SentiCircle to redefine the value of the word polarity in a lexicon opinion based on the aspect and determine the value of the polarity of the new opinion word. An overview of the methods we propose can be seen in Fig. 5.

a) Aspect *A* data

We take aspect *A* data from Aspect Categorization results. For example, if we want to determine the opinion lexicon for the Food quality aspect, the aspect *A* data is all data that categorized into the Food quality aspect in the Aspect Categorization stage.

b) Opinion Lexicon

In this study we used Lexicon Opinion from [7] which only gives negative or positive polarity, however it does not give the polarity values of the opinion words. From Fig. 6 in line 6, we assign positive opinion word with value 0.75, which is determined based on $\theta = 2.36\ radian$. In line 7, we assign negative opinion word with value -0.75, which is determined based on $\theta = -2.36\ radian$.

c) Opinion lexicon Polarity based on SentiCircle

In this part we redefine the polarity of the word opinion in the lexicon opinion and determine the polarity of the word new opinion. The determination of an opinion word polarity can be seen in Fig. 6.

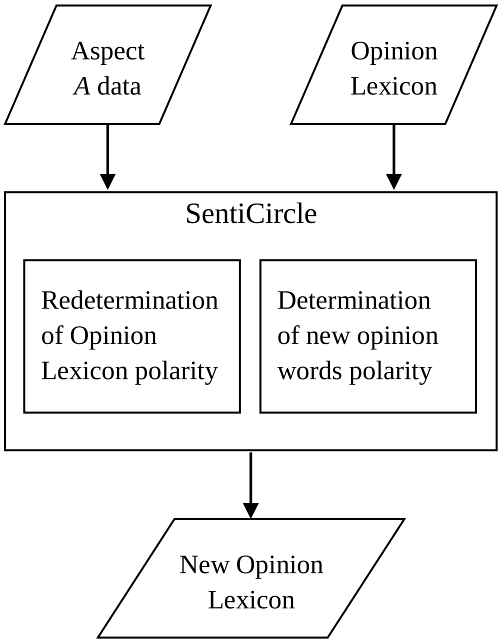

Figure. 5 HEOLS

Start
1. With Eqs. (4) and (5)
2. Set aspect A data as d
3. Set the opinion word that the polarity will be determined as m
4. Limit the c_i with Opinion Lexicon
5. With Eq. (6)
6. Set $Prior\_sentiment(c_i) = 0.75$ if c_i in positive opinion lexicon
7. Set $Prior\_sentiment(c_i) = -0.75$ if c_i in negative opinion lexicon
8. With Eqs. (7) and (8)
9. Determine the x_i and y_i position of c_i
10. With Eq. (9)
11. Calculate the m position
12. x_m and y_m = m position
13. If y_m > 0
14. Set m polarity = "positive"
15. Else
16. Set m polarity = "negative"

End

Figure. 6 Opinion lexicon polarity based on SentiCircle

Start
1. Score = 0
2. For each opinion word *ow* in sentence *s*
3. If *ow* in positive lexicon aspect *A*
4. Score = Score + 1
5. Else if *ow* in negative lexicon aspect *A*
6. Score = Score – 1

   Endif
7. Endfor

   If there is negation word around *ow*

   Score = Score * (-1)

   Endif
8. If Score > 0
9. Return "positive"
10. Else
11. Return "negative"
12. Endif

End

Figure. 7 Sentiment assignment

3.5.3. Sentiment assignment

The process of determining sentiments is determined by using the aspect A opinion lexicon from HEOLS results and a list of negation word. We use the same negation word used in [9]. The polarity assignment algorithm can be seen in Fig. 7.

3.5.4. Comparison

We compared our opinion words polarity from HEOLS with the opinion word polarity from: 1) Opinion Lexicon; 2) the first sense of adjective word SentiWordNet [30] (positive if the SentiWordNet score > 0 and vice versa), we use SentiWordNet because it was used in previous research [31, 32]; and 3) same as in point 2 but we add Word Sense Disambiguation (WSD) using Adapted Lesk [33] to improve the performance [34-36].

4. Results and analysis

This section explains the results of the experiments we have done. We evaluate the results using Precision (P), Recall (R) and F1measure (F) with Eqs. (12), (13), and (14), respectively.

$$P = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (12)$$

$$R = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (13)$$

$$F = 2 \times \frac{P \times R}{P + R} \quad (14)$$

Table 7. Aspect categorization results

Metode ATE	**Glove** [4, 5, 29]			**Word2vec** [29]			**Fasttext** [8]			**ELMo**			**Hybrid ELMo-Wikipedia**		
	P	***R***	***F***	***P***	***R***	***F***	***P***	***R***	***F***	***P***	***R***	***F***	***P***	***R***	***F***
DP	0.68	0.71	0.70	0.70	0.71	0.70	0.68	0.71	0.69	0.72	0.72	0.72	0.78	0.72	**0.75**
LDA	0.60	0.75	0.67	0.61	0.74	0.67	0.64	0.76	0.69	0.63	0.76	0.69	0.75	0.76	**0.75**
LDA2Vec	0.65	0.76	0.70	0.67	0.75	0.71	0.72	0.77	0.75	0.71	0.76	0.73	0.79	0.77	**0.78**
PLSA	0.66	0.73	0.69	0.70	0.73	0.71	0.72	0.74	0.73	0.71	0.74	0.72	0.80	0.74	**0.77**
DP*	0.67	0.79	0.72	0.69	0.79	0.74	0.66	0.79	0.72	0.70	0.81	0.75	0.76	0.81	**0.78**
LDA*	0.61	0.81	0.69	0.61	0.80	0.69	0.64	0.82	0.72	0.64	0.82	0.72	0.75	0.82	**0.78**
LDA2Vec*	0.65	0.81	0.72	0.68	0.80	0.73	0.71	0.82	0.76	0.71	0.80	0.75	0.78	0.81	**0.80**
PLSA*	0.65	0.80	0.71	0.69	0.80	0.74	0.70	0.80	0.75	0.71	0.80	0.75	0.79	0.81	**0.80**

Table 8. Comparison of aspect categorization results

Kalimat	**Aspect term extracted**	**Aspect categorization results**
I tend to judge a sushi restaurant by its sea urchin, which was heavenly at sushi rose.	restaurant	HEW = Food q. Glove = Service q. Word2vec = Food q. Fasttext = Service q. ELMo = Food q. **Label = Food**
The restaurant looks out over beautiful green lawns to the Hudson River and the Statue of Liberty.	restaurant	HEW = Physical e. Glove = Service q. Word2vec = Food q. Fasttext = Service q. ELMo = Food q. **Label = Physical e.**

Table 9. Sentiment analysis results

Methods	***P***	***R***	***F***
OL	0.90	0.69	0.78
HEOLS	0.85	0.83	**0.84**
S	0.78	0.53	0.63
S + WSD	0.78	0.54	0.64

4.1 Aspect categorization

The results of aspect categorization can be seen in Table 7. The method in the table that has a sign (*) is the method we modified. Based on the table it can be seen that the results of our modification f1measure is the best. It means that the modifications we made, namely the Noun Extraction and Aspect Term Filter can provide a more accurate prediction of the aspect categories. For example, the sentence "The sushi seemed pretty fresh and was adequately proportioned" with DP it does not produce any aspect predictions because there are no extracted aspects. With the method that we modified, we can extract the aspect terms in the sentence, "sushi" and with this aspect term we can predict the aspect category correctly, namely Food quality.

In Table 7 it can also be seen that the Hybrid ELMo and Wikipedia methods that we propose always have a higher f1measure compared to Glove, Word2vec, Fasttext and ELMo using any ATE method. This shows that the method we propose has the best results and can also be trusted. The cause of these results is that the vector results from HEW can recognize the context of the aspect term. For example, in Table 8, there are two different reviews with the same aspect term, namely "restaurant". The first review discusses the aspects of Food quality, the method of determining the vector with Word2vec, ELMo, and with our proposed method predict correctly. In the second review that discusses the aspects of Physical environment, only the HEW method that we proposed that can predict the aspect category correctly, the other four methods have the same predictions as the first review.

4.2 Sentiment analysis

The results of sentiment analysis can be seen in Table 9. From the table it can be seen that the Hybrid Expanded Opinion Lexicon-SentiCircle has a higher f1measure. That is because HEOLS can change the polarity value of the opinion word in the Opinion Lexicon based on aspects and can determine the polarity value of the new opinion word that is not in the Opinion Lexicon.

Examples of opinion word that the polarity change is the word "cheap". The word "cheap" both in Opinion Lexicon (OL) and SentiWordNet (S) is negative, whereas if it is in the aspect of Price fairness then “cheap” should be positive. By using the HEOLS we can change the polarity value of the word "cheap" which was originally negative to be

Received: August 5, 2019

positive in the aspect of Price fairness. It can be seen in Fig. 8 that the result of determining the polarity of the word opinion "cheap" (blue dot) is on the positive y axis so the polarity of the word "cheap" is positive. Therefore, the predicted results from the Sentiment Analysis with HEOLS in Table 10 are correct.

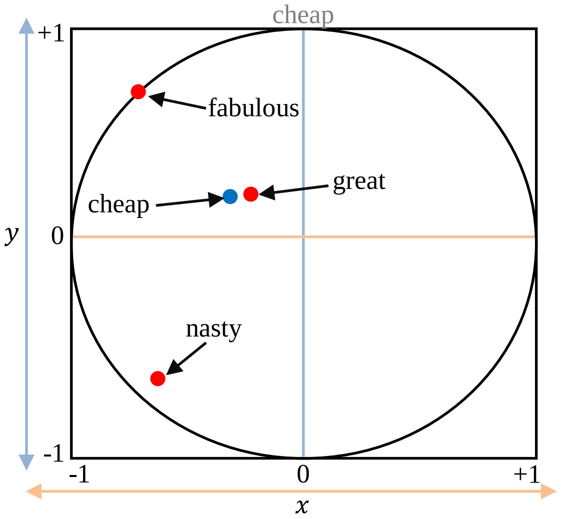

Figure. 8 Polarity of "cheap"

Table 11. Comparison of sentiment analysis in Price fairness aspect

Review	**Opinion extracted**	**Aspect categorization results**
The prices were cheap compared to the quality of service and food.	cheap	HEOLS = Positive OL = Negative S = Negative S + WSD = Negative **Label = Positive**
Their prices are so cheap!	cheap	HEOLS = Positive OL = Negative S = Negative S + WSD = Negative **Label = Positive**

Table 12. Comparison of sentiment analysis in physical environment aspect

Review	**Opinion extracted**	**Aspect categorization results**
The atmosphere is noisy and the waiters are literally walking around doing things as fast as they can.	noisy	HEOLS = Positive OL = Negative S = Negative S + WSD = Negative **Label = Negative**

Table 13. Aspect based sentiment analysis results

Aspect	**Sentiment**	**Amount (%)**
Physical environment	Positive	16.05
	Negative	5.97
Food quality	Positive	27.16
	Negative	9.26
Service quality	Positive	18.31
	Negative	6.79
Price Fairness	Positive	9.26
	Negative	7.20
Total		100.00

Examples for the results of determining the polarity for new opinion words are the words "tasty" and "yummy". These two words are not in the Opinion Lexicon, so the results of the SA are always negative. Hybrid Expanded Opinion Lexicon SentiCircle that we propose gives positive polarity for both words so that the results of SA in the sentence "food was really tasty" and "the pizza is yummy" are positive according to the label.

Considering the results of Sentiment Analysis in Table 9, the value of precision from HEOLS is lower than Opinion Lexicon. This is caused by not all of the HEOLS results are correct; there are still errors in determining the polarity of the opinion words. For example, Physical environment aspect in Table 11, the polarity value of the word "noisy" in Opinion Lexicon is negative but HEOLS provides positive polarity. Even though the polarity of the word “noisy” should not change and still has a negative value. That causes the results of the HEOLS to be wrong, so that the precision is decreased.

4.3 Aspect based sentiment analysis

The results of Aspect based Sentiment Analysis in Table 12 show that the restaurant has a positive sentiment in every aspect.

5. Conclusion

This study proposes two hybrid methods for Aspect based Sentiment Analysis in restaurant reviews and modification of Double Propagation (DP) and Topic Modelling (TM) methods. Our modification of DP and TM methods can increase the f1measure of Aspect Categorization by average of 2%. Aspect Categorization using our proposed method Hybrid ELMo-Wikipedia has better f1measure results than using similar methods (Glove, Word2vec, Fasttext and ELMo) with any method used in ATE by average 6%. Then the Hybrid

 DOI: 10.22266/ijies2019.1231.05

Received: August 5, 2019

Expanded Opinion Lexicon-SentiCircle that we propose can change the polarity value of the Opinion Lexicon according to its aspects and can determine the polarity value of the new opinion words. It can be seen that the word "cheap" in the Opinion Lexicon was initially negative, with the method we propose the word "cheap" to be positive in the aspect of Price fairness. Then the examples of new words that are not in the Opinion Lexicon are the words "yummy" and "tasty", in both words the method we propose gives a positive polarity value. Lastly, Sentiment Analysis results with our proposed Hybrid Expanded Opinion Lexicon SentiCircle can increase the f1measure by 6%. For future work, the redetermination of word polarity in the Opinion Lexicon can be improved by reducing error i.e. the word "noisy" redetermined as positive with our proposed method, although it should remain negative (unchanged).

Acknowledgments

The authors would like to thank to Institut Teknologi Sepuluh Nopember, Direktorat Riset dan Pengabdian Masyarakat, Direktorat Jenderal Penguatan Riset dan Pengembangan, the Ministry of Research, Technology, and Higher Education of Indonesia for financing the research.

References

[1] A. García-Pablos, M. Cuadros, and G. Rigau, "W2VLDA: Almost unsupervised system for Aspect Based Sentiment Analysis", *Expert Systems with Applications*, Vol. 91, pp. 127–137, 2018.

[2] S. Gagić, D. Tešanović, and A. Jovičić, "The Vital Components of Restaurant Quality that Affect Guest Satisfaction", *Turizam*, Vol. 17, No. 4, pp. 166–176, 2013.

[3] 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.

[4] R. A. 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.

[5] D. A. K. 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] N. Akhtar, N. Zubair, A. Kumar, and T. Ahmad, "Aspect based Sentiment Oriented Summarization of Hotel Reviews", *Procedia Computer Science*, Vol. 115, pp. 563–571, 2017.

[7] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews", In: *Proc. of International Conf. on Knowledge Discovery and Data Mining*, pp. 169–177, 2004.

[8] A. Firmanto and R. Sarno, "Aspect-Based Sentiment Analysis Using Grammatical Rules, Word Similarity and SentiCircle", *International Journal of Intelligent Engineering and Systems*, Vol. 12, No. 5, pp. 190–201, 2019.

[9] 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.

[10] H. Liu, "Dependency Distance as a Metric of Language Comprehension Difficulty", *Journal of Cognitive Science*, Vol. 9, pp. 159–191, 2008.

[11] Q. Liu, Z. Gao, B. Liu, and Y. Zhang, "Automated rule selection for opinion target extraction", *Knowledge-Based Systems*, Vol. 104, pp. 74–88, 2016.

[12] H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li, and L. Zhao, "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey", *Multimedia Tools and Applications*, Vol. 78, No. 11, pp. 15169–15211, 2019.

[13] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation", *Journal of Machine Learning Research*, Vol. 3, pp. 993–1022, 2003.

[14] T. Hofmann, "Probabilistic Latent Semantic Analysis", In: *Proc. of the 15th Conf. on Uncertainty in Artificial Intelligence*, pp. 289–296, 1999.

[15] C. E. Moody, "Mixing Dirichlet Topic Models and Word Embeddings to Make lda2vec", *CoRR*, 2016.

[16] R. Lourdusamy and S. Abraham, "A Survey on Text Pre-processing Techniques and Tools", *International Journal of Computer Sciences and Engineering*, Vol. 6, No. 3, pp. 148–157, 2019.

[17] 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 Conf. on Information and*

Communication Technology, pp. 1–6, 2016.

[18] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations", In: *Proc. of North American Association for Computational Linguistics*, pp. 2227–2237, 2018.

[19] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global Vectors for Word Representation", In: *Proc. of the 2014 Conf. on Empirical Methods in Natural Language Processing*, pp. 1532–1543, 2014.

[20] 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, 2016.

[21] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space", In: *Proc. of the International Conf. on Learning Representations*, pp. 1–12, 2013.

[22] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: A System for Large-Scale Machine Learning", In: *Proc. of the 12th USENIX Conf. on Operating Systems Design and Implementation*, pp. 265–283, 2016.

[23] J. Martinez-Gil, "An overview of textual semantic similarity measures based on web intelligence", *Artificial Intelligence Review*, Vol. 42, No. 4, pp. 935–943, 2012.

[24] P. Xia, L. Zhang, and F. Li, "Learning similarity with cosine similarity ensemble", *Information Sciences*, Vol. 307, pp. 39–52, 2015.

[25] L. Derczynski, D. Maynard, G. Rizzo, M. Van Erp, G. Gorrell, R. Troncy, J. Petrak, and K. Bontcheva, "Analysis of named entity recognition and linking for tweets", *Information Processing and Management*, Vol. 51, No. 2, pp. 32–49, 2015.

[26] J. R. Finkel, T. Grenager, and C. Manning, "Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling", In: *Proc. of the 43rd Annual Meeting on Association for Computational Linguistics*, pp. 363–370, 2007.

[27] M. Pontiki, D. Galanis, H. Papageorgiou, and I. Andrutsopoulos, "SemEval-2016 Task 5: Aspect Based Sentiment Analysis", In: *Proc. of the 10th International Workshop on Semantic Evaluation*, pp. 19–30, 2016.

[28] A. X. Chang and C. D. Manning, "TokensRegex: Defining cascaded regular expressions over tokens", 2014.

[29] D. H. Pham and A. C. Le, "Exploiting multiple word embeddings and one-hot character vectors for aspect-based sentiment analysis", *International Journal of Approximate Reasoning*, Vol. 103, pp. 1–10, 2018.

[30] S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining", In: *Proc. of the 7th International Conf. on Language Resources and Evaluation*, pp. 2200–2204, 2010.

[31] 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.

[32] C. Diamantini, A. Mircoli, D. Potena, and E. Storti, "Social information discovery enhanced by sentiment analysis techniques", *Future Generation Computer Systems*, Vol. 95, pp. 816–828, 2019.

[33] S. Banerjee and T. Pedersen, "An Adapted Lesk Algorithm for Word Sense Disambiguation Using WordNet", In: *Proc. of the 3rd International Conf. on Computational Linguistics and Intelligent Text Processing*, pp. 136–145, 2007.

[34] 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.

[35] 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.

[36] C. Hung and S. J. Chen, "Word sense disambiguation based sentiment lexicons for sentiment classification", *Knowledge-Based Systems*, Vol. 110, pp. 224–232, 2016.