Aspect-based Sentiment Analysis for Indonesian Restaurant Reviews

Devina Ekawati
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
devina.ekawati@gmail.com

Masayu Leylia Khodra
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
masayu@stei.itb.ac.id

***Abstract*—Aspect-based sentiment analysis summarizes what people like and dislike from reviews of products or services. In this paper, we adapt the first rank research at SemEval 2016 to improve the performance of aspect-based sentiment analysis for Indonesian restaurant reviews. We use six steps for aspect-based sentiment analysis i.e.: preprocess the reviews, aspect extraction, aspect categorization, sentiment classification, opinion structure generation, and rating calculation. We collect 992 sentences for experiment and 383 sentences for evaluation. We conduct experiment to find best feature combination for aspect extraction, aspect categorization, and sentiment classification. The aspect extraction, aspect categorization, and sentiment classification have F1-measure value of 0.793, 0.823, and 0.642 respectively.**

Keywords—aspect-based sentiment analysis; aspect extraction; aspect categorization; sentiment classification; rating calculation

I. Introduction

User-generated review sites, like Amazon [1] and TripAdvisor [2], provide subjective opinions about a product or service. Those sites have large number of reviews so it is hard for human to summarize people opinions. Thus, automated sentiment analysis is needed. Results of sentiment analysis may help consumers to decide whether they will buy the products or using the services. Manufacturers can use it to improve their product or service qualities. Based on the level of sentiment analysis (i.e. document level, sentence level, and aspect level [3]), sentiment analysis on aspect level gives better results as it more describes what reviewers like and dislike. Therefore, we use aspect-based sentiment analysis (ABSA) on this paper.

One research that has shown success in ABSA is research conducted by [4] which is the first rank research in International Workshop on Semantic Evaluation (SemEval) 2016 for text level subtask. Hercig et al. [4] use lexical features, syntactic features, and distributional semantic model. F1-measure of aspect extraction, aspect categorization, and sentiment classification for restaurant reviews are 0.671, 0.682, and 0.817 respectively.

Hercig et al.'s [4] method has not been applied in Indonesian text. This paper aims to generate a corpus of Indonesian reviews using the same approach in SemEval 2016. Then, [4]'s method is applied on the corpus to improve performance of ABSA in Indonesian conducted by [5] which has F1-measure of 0.554 for overall system. Unlike [5], we use all sentences (objective and subjective sentence) for ABSA. The generated opinion structures are also different. The generated opinion structures in this paper consist of aspect, category of aspect, and sentiment polarity, while the opinion structures generated by [5] consist of aspect, category of aspect, opinion word, and sentiment of opinion word.

Since Hercig et al. [4] build models independently for three subtasks and cannot be applied directly to ABSA, this paper combines the results of those models to generate opinion structures which are used to calculate rating. This paper does not use all lexical and syntactic features as in [4] because some feature extraction tools are unavailable for Indonesian text and some features are not applicable for dataset used in this paper.

This paper is organized in the following manner. Section 2 discusses related work. We describe ABSA method we used in section 3. Section 4 shows the experimental results, evaluation results, and analysis of the evaluation results. The conclusion as well as future plans for this work are in section 5.

II. Related Work

There are six steps in aspect-based sentiment analysis: entity extraction and categorization, aspect extraction and categorization, sentiment polarity classification, time extraction, opinion holder extraction, and opinion structures generation [3]. Entity, time, and opinion holder extraction is a problem of named entity recognition (NER). Supervised learning, such as Conditional Random Field (CRF) [6], is commonly used to solve NER problem.

Supervised learning is also a common approach for aspect extraction. Hercig et al. [4] employed CRF with lexical and syntactic features, such as token, POS tag, dependency path, and features from Continuous Bag-of-Word (CBOW), which is one of distributional semantic model [7]. Distributional semantic model is based on an assumption that the meaning of a word can be inferred from it usages [8].

Lexicon-based approach [9] and SentiWordNet-based approach [10] can be used for sentiment classification. Hercig et al. [4] employed supervised learning to classify sentiment of categories in a sentence. They use Maximum Entropy (MaxEnt) classifier [11] with lexical features, syntactic features, and features from CBOW and Global Vectors for word representation (GloVe) [12].

Gojali and Khodra [5] merged aspect and sentiment extraction into one step by employing CRF with token and POS

978-1-5386-3001-3/17/$31.00 ©2017 IEEE

tag as features. Each token is assigned into one of four entities: aspect, positive opinion, negative opinion, and other. After labelling each token, aspect is paired by its matching opinion. Then, the orientation or polarity of the opinion words are determined by checking the presence of negation word with five words distance from the opinion words.

After aspect extraction, the aspects can be categorized into categories. One approach is to use WordNet. Liu et al. [13] use WordNet to find synonym between aspect. For example, photo, picture, and image can be categorized into one category because those three words are synonym in WordNet. Gojali and Khodra [5] also use WordNet to find similarity between aspect and predefined seed word for restaurant categories. Hercig et al. [4] use supervised learning using multilabel classification with MaxEnt algorithm. They also use lexical and syntactic features as in aspect extraction and sentiment classification step and features from CBOW, GloVe, and Latent Dirichlet Allocation (LDA) [14].

III. THE PROPOSED METHOD

As stated before, this paper adapts ABSA method of Hercig et al. [4] to improve the performance of aspect-based sentiment analysis in Indonesian. In our work, there are six steps to do ABSA which is shown in Fig 1.

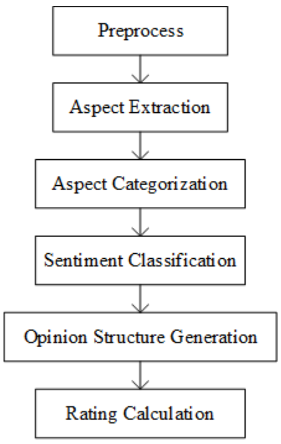

Fig 1. Architecture of the Application

Preprocess step includes sentences splitting, sentence normalization, and stop word elimination. InaNLP [15] is used to do the preprocess of the reviews. Results of preprocess step is used to build distributional semantics model, aspect extraction, aspect categorization, and sentiment classification. There are three distributional semantic models: CBOW model which is a model of word2vec, GloVe, and LDA. We build those three models by using 11470 reviews which consist of 36766 sentences. We use [16] to build CBOW and LDA model. GloVe implemented by [12] is used to build GloVe model. We employ k-means algorithm for clustering distributional semantic model implemented in [17].

For aspect extraction step, we employ CRF algorithm to get label for each token with CRFSuite [18]. We label each token in a sentence with three labels: ASPECT-B, ASPECT-I, and OTHER. Features for aspect extraction is shown in TABLE 1. Feature extraction is conducted for each token with five words context window centered at currently processed word as in [4]. We use CBOW for clusters and clusters bigram. N value of 1 to 3 is used for N-gram feature. POS tag and head words are obtained with SyntaxNet [19].

TABLE 1. FEATURES FOR ASPECT EXTRACTION

Feature	Definition
Bag of N-gram	The occurrence of a n-gram in the context window
Bag of POS N-gram	The occurrence of POS n-gram in the context window
Clusters	The occurrence of word's cluster at a given position
Clusters Bigram	The occurrence of word's cluster bigram at a given position
Head Words	Word that determines the syntactic category of that word from the dependency tree
Learned Target Dictionary	Presence of a word from training data of aspect term

Since SemEval 2016 has independent tasks for aspect extraction and aspect categorization, we need to add sentence preprocess before applying aspect categorization. We split a sentence into two sentences if it has contrary conjunction, such as *tetapi* (but), *namun* (however), etc. This preprocess solves problem when a sentence has two or more aspects with same category but different sentiment polarity. We define two rules i.e.:

1. If both clauses before and after contrary conjunction contain extracted aspects, we split the sentence into two sentences by using contrary conjunction as delimiter. For example, "*saya suka asparagusnya* ***namun*** *salmonnya tidak suka"* (I like the asparagus **but** I don't like the salmon) is splitted into "*saya suka asparagusnya"* (I like the asparagus) and "*namun salmonnya tidak suka* (but I don't like the salmon)".
2. If clause before contrary conjunction contains aspect but clause after contrary conjunction does not contain aspect, we also split it into two sentences by using contrary conjunction as delimiter and inserting aspect from the first sentence into the second sentence. For example, "*harga makanannya mahal* ***namun*** *sebanding*" (food price is expensive **but** comparable)" is splitted into "*harga makanannya mahal*" (food price is expensive) and "*harga makanannya sebanding*" (food price is comparable).

We build multilabel binary relevance classifier with MaxEnt algorithm for aspect categorization. The illustration of binary relevance classifier is shown in Fig 2. As in [5], we define four categories: food, place, price, and service. Each category has its binary classifier so the total classifier is same as the total of category. Each sentence in the corpus is labeled with Boolean value for each category: true if a sentence has certain category and false otherwise. The classifier for each category classify the Boolean value. After that, we collect categories with true values as multilabel output.

Since Hercig et al. [4] define lexical and syntactic features of this step for three languages (Chinese, English, and Spanish), we select general features of those three languages. Features that we use for this step is shown at TABLE 2. Bag of clusters is obtained from CBOW, LDA, and GloVe. We use N value of 1 and 2 for bag of N-gram. We employ [17] to build aspect categorization model and also predict the categories of the sentences.

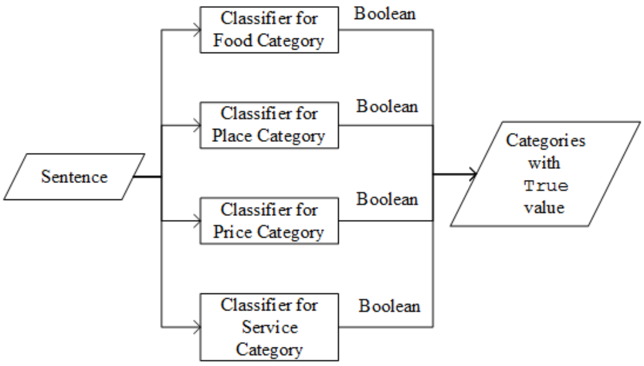

Fig 2. Binary Relevance Classifier for aspect categorization

TABLE 2. GENERAL FEATURES FOR ASPECT CATEGORIZATION

Feature	Definition
Bag of N-grams	The occurrence of a N-gram in the context window
Bag of Clusters	The occurrence of word's cluster in the context window

We apply the same algorithm as in aspect categorization for sentiment classification. As in [5], we define two labels of sentiment for each category in a sentence: positive and negative. For example, a sentence has food and place category. Each category has its own classifier to classify sentiment of categories in the sentence. To classify the sentiment polarity, we use classifier for food and place category. As in aspect categorization, we use general features from Chinese, French, English, and Spanish. The features are shown in TABLE 3. Bag of clusters is obtained from CBOW, LDA, and GloVe. We use 2 to 5 words skips for Skip-bigram [4]. N value of 1 and 2 is also used for bag of N-gram feature.

TABLE 3. FEATURES FOR SENTIMENT CLASSIFICATION

Feature	Definition
Bag of N-grams	The occurrence of a N-gram in the context window
Bag of Head Words	Bag of word that determines the syntactic category of that word from the dependency tree
Bag of Clusters	The occurrence of word's cluster in the context window
Bag of k-skip-bigram	Bag of N-gram which has skipped over gaps.

For opinion structure generation, we employ CBOW model to find similarity between extracted aspect with seed words for each category. Category that has maximum similarity score will be paired with the extracted aspect. For example, a sentence has food and place category and we want to pair "cake" with a category from the sentence. First, we find similarity score for each seed words for food and place category. The maximum similarity score for food and place category are 0.2 and 0.1 respectively. After that, we can pair "cake" with food category because similarity score for food category is higher than place category. The generated opinion structures are used to calculate *rating* for each category with equation as in [20].

IV. EXPERIMENT, EVALUATION, AND DISCUSSION

We collect 992 sentences of restaurant reviews from TripAdvisor for dataset [5]. As mentioned above, we need to split the sentences with contrary conjunctions before aspect categorization and sentiment classification. Therefore, the total sentences for aspect categorization and sentiment classification after sentence splitting are 1100. We use additional 485 sentences to build sentiment classifier because total sentence in training data for service and price category are much less than food and place category. We conduct experiment using one factor at a time strategy [21]. We observe controllable factors i.e. cluster number (100, 500, 1000, and 5000) and topics value (50, 100, 200, 300, 400, and 500) for LDA as in [4]. The experiment is performed independently for each step of ABSA using 10-fold cross validation. Results from previous step of ABSA is not propagated into the following step in experiment.

We also collect test data from TripAdvisor restaurant reviews to evaluate our model. The test data consists of 383 sentences [5]. Total tokens of ASPECT-B, ASPECT-I, and OTHER in data test are 352, 95, and 4459 respectively. Meanwhile, total labels of food, service, price, and place category are 104, 21, 29, and 134 respectively. For evaluation, we use precision, recall, and F1-measure. The results from aspect extraction used for sentence splitting for aspect categorization. Sentiment classification classify the sentiment polarity of each category in a sentence resulting from aspect categorization step.

A. *Aspect Extraction*

TABLE 4 and Fig 3 show token distribution and example of training data for aspect extraction step. For clusters and clusters bigram feature, there are four scenarios to experiment with the number of clusters. The best combination for this step is bag of N-gram, bag of POS N-gram, clusters with 5000 clusters, and clusters bigram with 100 clusters CBOW. The eight best results of experiment are shown at TABLE 5.

TABLE 4. TOKEN DISTRIBUTION OF ASPECT EXTRACTION IN TRAINING DATA

Label	Total Token
ASPECT-B	1353
ASPECT-I	435
OTHER	10667
Total	**12455**

```
Steak <ASPECT-B> salmon <ASPECT-I> dan <OTHER> sauce
<ASPECT-B> enak <OTHER>

Steak <ASPECT-B> salmon <ASPECT-I> and <OTHER> sauce
<OTHER> are <OTHER> delicious <OTHER>
```

Fig 3. Example of Training Data for Aspect Extraction

The evaluation results of aspect extraction (TABLE 6) show quite high F1-measure for ASPECT-B and OTHER label. However, F1-measure for ASPECT-I label is low because most tokens labeled as OTHER while total tokens with ASPECT-B and ASPECT-I label are much less compared to total tokens of OTHER label.

TABLE 5. BEST EXPERIMENTAL RESULTS OF ASPECT EXTRACTION

No	Features	Parameters	P	R	F
1	Learned Target Dictionary	-	0.977	0.665	0.741
2	Head word	Stop word elimination	0.682	0.491	0.526
3	Bag of N-gram	-	0.897	0.637	0.712
4	Bag of POS N-gram	-	0.732	0.648	0.679
5	Clusters	5000 clusters	0.873	0.691	0.755
6	Clusters Bigram	100 clusters	0.794	0.539	0.602
7	Bag of N-gram Bag of POS N-gram Clusters Clusters Bigram	- - 5000 clusters 100 clusters	**0.849**	**0.747**	**0.790**
8	Bag of N-gram Bag of POS N-gram Clusters Clusters Bigram Head Word	- - 5000 clusters 100 clusters -	0.842	0.749	0.788

TABLE 6. EVALUATION RESULTS OF ASPECT EXTRACTION

Label	Precision	Recall	F1
ASPECT-B	0.802	0.784	0.793
ASPECT-I	0.567	0.653	0.608
OTHER	0.987	0.977	0.978
Average	**0.783**	**0.804**	**0.793**

Our aspect extraction model can extract several aspects that do not appear in the training data, such as "*guramenya"* (the carp) and "*jus strawberry"* (strawberry juice). Because the model use bag of N-gram features, classification of a token is very dependent with the co-occurrence of word, POS tag, and clusters CBOW in the context window. One example of misclassification occurs in the sentence "*baik secara rasa dan tempat yang nyaman untuk bersantai"* (Both in taste and comfortable place to relax). The actual aspects for the sentence are "*rasa"* (taste) and "*tempat"* (place). But the extraction result is "*tempat"* (place). The model fails to extract "*rasa"* (taste) as aspect because "*rasa"* (taste) usually co-occurs with words related to food, such as "*enak"* (delicious) and "*makanan"* (food) in the training data. But the sentence does not have words related to food so misclassification occurs.

Another misclassification occurs in the sentence "*paling enak nasi gorengnya, sate ayam, steak braga permai, dan chicken snitzel"* (the most delicious foods are fried rice, chicken satay, braga permai steak, and chicken schnitzel). The expected aspects of the sentence are "*nasi gorengnya"* (fried rice), "*sate ayam"* (chicken satay), "*steak braga permai"* (braga permai steak), and "*chicken snitzel"* (chicken schnitzel). But the result of aspect extraction is "*nasi gorengnya"* (the fried rice). "*Sate ayam"* (chicken satay) and "*chicken snitzel"* (chicken schnitzel) are not extracted as aspect because "*sate"* (satay) and "*snitzel"* (schnitzel) never appear in the training data. Although "*steak*" and "*braga permai*" appear in training data, "*steak*" and "*braga permai*" never co-occur in training data within the context window. Besides that, "*braga permai*" appear in training data as a name of a restaurant, not a part of food name so it is never labeled as an aspect.

B. Aspect Categorization

The label distribution in training data and example of training data for aspect categorization are shown at TABLE 7 and TABLE 8 respectively. For bag of clusters feature, we also use four scenarios and additional six scenarios to experiment with number of topic for LDA. The best feature for this step is bag of clusters using CBOW model with 1000 clusters. Eight best experimental results of aspect categorization are shown at TABLE 9.

TABLE 7. LABEL DISTRIBUTION OF ASPECT CATEGORIZATION IN TRAINING DATA

Label	Total Sentence
Food	503
Service	97
Price	125
Place	440

TABLE 8. EXAMPLE OF TRAINING DATA FOR ASPECT CATEGORIZATION

Sentence	Food	Service	Price	Place
Saya suka makanannya (I like the food)	Yes	No	No	No
Tempat tetap nyaman, masakan tetap lezat, dan pelayanan tetap ramah (Place is still comfortable, cuisine is still delicious, and service is still friendly)	Yes	Yes	No	Yes

TABLE 9. BEST EXPERIMENTAL RESULTS OF ASPECT CATEGORIZATION

No	Features	Parameters	P	R	F1
1	Bag of N-gram	-	0.949	0.755	0.816
2	Bag of Clusters CBOW	1000 clusters	**0.943**	**0.805**	**0.856**
3	Bag of Clusters GloVe	5000 clusters	0.950	0.754	0.816
4	Bag of Clusters LDA	100 topics	0.950	0.764	0.826
5	Bag of N-gram Bag of Clusters CBOW Bag of Clusters GloVe	- 1000 clusters 5000 clusters	0.849	0.747	0.790
6	Bag of N-gram Bag of Clusters CBOW Bag of Clusters GloVe Bag of Clusters LDA	- 1000 clusters 5000 clusters 100 topics	0.950	0.794	0.849
7	Bag of N-gram Bag of Clusters CBOW Bag of Clusters LDA	- 1000 clusters 100 topics	0.951	0.793	0.849
8	Bag of N-gram Bag of Clusters CBOW	- 1000 clusters	0.956	0.796	0.852

TABLE 10 shows evaluation results of aspect categorization. The F1-measure of price category is very high while F1-measure of food, price, and place category is quite high. The misclassification for price category never occurs because all sentences which have price category always have "*harga* (price)" in the test data.

TABLE 10. EVALUATION RESULTS OF ASPECT CATEGORIZATION

Category	Precision	Recall	F1
Food	0.658	0.906	0.763
Service	0.739	0.773	0.756
Price	1.000	1.000	1.000
Place	0.705	0.860	0.775
Average	**0.775**	**0.884**	**0.823**

Some misclassifications occur when the sentence has word "*restoran"* (restaurant) or restaurant name. For example, sentence "*kami datang merayakan hari jadi kami ke Atmosphere café dengan harapan yang cukup tinggi"* (we come to Atmosphere café for celebrating our anniversary with high expectation)" is classified as place category while the sentence actually does not have any category. The sentence that have word "*atmosphere*" only appear once in the training data and it is labeled as place category so the sentence misclassified as place category. Another misclassification happens in sentence "*ini restoran jaman dulu yang sampai sekarang eksis"* (this is an old restaurant that still exists until now). The sentence has word "*restoran"* (restaurant) so it is classified as place category by the model. But word "*restoran"* (restaurant) does not have any sentiment so the sentence is not labeled as place category even though it has word "*restoran"* (restaurant).

Misclassification can happen if words never co-occur in the training data as in aspect extraction. For example, sentence "*tetapi yang spesial lingkungannya"* (but the speciality is the environment) is classified as food category while the sentence actually has place category. Most sentences have word "*spesial"* (special) labeled as food category in the training data because it co-occurs with word related to food and it never co-occurs with word "*lingkungannya"* (environment). Besides that, the word "*lingkungannya"* (environment) never appears in training data. Because of that, the sentence is misclassified as food category.

C. Sentiment Classification

TABLE 11 and TABLE 12 show the sentiment distribution of each restaurant category for sentiment classification and example of training data respectively.

TABLE 11. LABEL DISTRIBUTION OF SENTIMENT CLASSIFICATION IN TRAINING DATA

Category	Label	Total Sentence
Food	Positive	493
	Negative	61
Service	Positive	239
	Negative	54
Price	Positive	231
	Negative	145
Place	Positive	401
	Negative	63

TABLE 12. EXAMPLE OF TRAINING DATA FOR SENTIMENT CLASSIFICATION

Sentence	Food	Service	Price	Place
Saya suka makanannya (I like the food)	Positive	-	-	-
Tempat tetap nyaman, masakan tetap lezat, dan pelayanan tetap ramah (Place is still comfortable, cuisine is still delicious, and service is still friendly)	Positive	Positive	-	Positive

We use four scenarios for bag of clusters feature to experiment with number of clusters. The best performance is achieved by using three features: bag of N-gram, clusters with CBOW model and GloVe model. We use 500 clusters for CBOW and Glove clusters. Eight best experimental results are shown at TABLE 13.

TABLE 14 shows evaluation results of sentiment classification for each restaurant category. The F1-measure is quite low for food, service, and place category because misclassification that occurred in aspect categorization step.

TABLE 13. BEST EXPERIMENTAL RESULTS OF SENTIMENT CLASSIFICATION

No	Features	Parameters	P	R	F1
1	Bag of N-gram	-	0.758	0.666	0.666
2	Bag of Clusters CBOW	1000 clusters	0.831	0.738	0.748
3	Bag of Clusters GloVe	5000 clusters	0.842	0.739	0.752
4	Bag of Head Words	-	0.481	0.741	0.459
5	Bag of k-skip-bigram	-	0.399	0.500	0.438
6	Bag of N-gram Bag of Clusters CBOW Bag of Clusters GloVe	- 1000 clusters 5000 clusters	**0.867**	**0.772**	**0.786**
7	Bag of N-gram Bag of Clusters CBOW Bag of Clusters GloVe Bag of k-skip-bigram	- 1000 clusters 5000 clusters 100 topics	0.867	0.772	0.786
8	Bag of N-gram Bag of Clusters CBOW Bag of Clusters GloVe Bag of Head Words	- 1000 clusters 5000 topics -	0.797	0.718	0.723

TABLE 14. EVALUATION RESULTS OF SENTIMENT CLASSIFICATION

Category	Precision	Recall	F1
Food	0.361	0.538	0.427
Service	0.588	0.643	0.614
Price	0.950	0.909	0.924
Place	0.619	0.627	0.604
Average	**0.629**	**0.679**	**0.642**

One example of misclassification occurs in the sentence "*pernah order gurame bakarnya, guramenya sedikit overburn*" (order grilled carp, it slightly overburnt). The sentence has negative sentiment because of word "*overburn*" (overburnt). But it is misclassified as positive because the word never appears on training data. Another misclassification occurs in sentence "*tempatnya mahal*" (place is expensive) because of out of vocabulary opinion word. Opinion word "*mahal*" (expensive) never appears in training data for place category so it is misclassified as positive. Opinion word "*mahal*" (expensive) only appears in training data for food and price category.

Another misclassification occurs when the sentence has multiple sentiment for one category, as in sentence "*tempatnya cukup enak hanya terlalu gelap dan kurang privasi karena semua customer berada di ruangan yang sama dan jarak antar mejanya cukup dekat*" (the place is good but it's too dark and not private because all customers are in one room and the space between tables is close enough). The sentence supposed to be split by sentence splitting rule as explained before because place category has two different sentiment: positive because of opinion word "*enak* (good)" and negative because of opinion word "*gelap* (dark). But the rules fail because word "*hanya*" is not a contrary conjunction in Indonesian.

D. Opinion Structures Generation and Rating Calculation

After aspect extraction, aspect categorization, and sentiment classification, the opinion structures are generated to calculate rating. The opinion structures and rating for each category of a restaurant review (Fig 4) is shown at TABLE 15. As stated before, the opinion structures consist of aspect, aspect category, and sentiment polarity. Rating is calculated for each category using formula as in [20]. For example, category place in has two positive aspects and zero negative aspect so the rating for place category is

$$Rating = \left(\frac{P}{P+N} \times 4\right) + 1 = \left(\frac{2}{2+0} \times 4\right) + 1 = 5.00$$

```
tempat nya nyaman, pemandangan bagus, harga
terjangkau. untuk makanannya menurut saya enak,
tapi minuman cocktail nya kurang enak,
bartendernya masih perlu belajar lagi sepertinya.

(The place was comfortable, the view was nice, and
the price was affordable. In my opinion, the food
was good, but the cocktail was not too good, the
bartender still has a lot to learn.)
```

Fig 4 Example of Restaurant Review

TABLE 15. EXAMPLE OF GENERATED OPINION STRUCTURES AND RATING

Category	Sentiment	Aspect	Rating
Food	Positive	*Makanannya* (food)	5.00
	Negative	-	
Place	Positive	-	0.00
	Negative	-	
Price	Positive	*Harga* (price)	5.00
	Negative	-	
Service	Positive	*Tempat* (place) *Pemandangan* (view)	5.00
	Negative	-	

V. CONCLUSION

Aspect-based sentiment analysis has six steps i.e.: preprocess, aspect extraction, aspect categorization, sentiment classification, opinion structure generation, and rating calculation. For aspect extraction, CRF is employed with four features: bag of N-gram, bag of POS N-gram, clusters CBOW (5000 clusters) and clusters bigram CBOW (100 clusters) with F1-measure of 0.793. We build MaxEnt classifier for aspect categorization and sentiment classification. Feature used for aspect categorization is bag of clusters CBOW (1000 clusters) with F1-measure of 0.823. For sentiment classification, we employ bag of N-gram, bag of clusters CBOW (1000 clusters), and bag of clusters GloVe (5000 clusters) with F1-measure of 0.642.

For improving the performance, additional restaurant reviews can be added to build distributional semantic model. Adding training data for aspect extraction, aspect categorization, and sentiment classification step can also improve the performance of the models. Besides that, we need to improve preprocess step because there are some informal words like “untk” and “bgtu” which should be normalize into “*untuk* (for)” and “*begitu* (like that)” but those two words are not normalized.

ACKNOWLEDGEMENT

We thank Ahmad Naufal Farhan for providing restaurant reviews dataset from TripAdvisor to build distributional semantic model.

REFERENCES

[1] "Amazon," [Online]. Available: https://www.amazon.com/.

[2] "TripAdvisor," [Online]. Available: https://www.tripadvisor.co.id/.

[3] B. Liu, Sentiment Analysis and Opinion Mining, 2012.

[4] T. Hercig, T. Brychc, L. Svoboda and M. Konko, "UWB at SemEval-2016 Task 5: Aspect Based Sentiment Analysis," in *The 8th International Workshop on Semantic Evaluation*, San Diego, 2016.

[5] S. Gojali and M. L. Khodra, "Aspect Based Sentiment Analysis for Review Rating Prediction," in *ICAICTA*, 2016.

[6] J. Lafferty, A. McCallum and F. C. N. Pereira, "Conditional Random Fields: Probabilistic Models," in *Proceedings of International Conference on Machine Learning*, 2001.

[7] T. Mikolov, K. Chen, G. Corrado and J. Dean, "Efficient estimation of word representations in vector space," 2013. [Online]. Available: https://arxiv.org/pdf/1301.3781v3.pdf. [Accessed 12 November 2016].

[8] S. Evert, "Distributional semantic models," in *Proceedings of Human Language Technologies: The 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Los Angeles, 2010.

[9] X. Ding, B. Liu and P. S. Yu, "A holistic lexicon-based approach to opinion mining," in *Proceedings of the Conference on Web Search and Web Data Mining (WSDM-2008)*, 2008.

[10] S. Baccianella, A. Esuli and F. Sebastiani, "SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," in *Proceedings of Seventh Conference on International Language Resources and Evaluation (LREC’10)*, 2010.

[11] A. L. Berger, S. A. D. Pietra and V. J. D. Pietra, "A Maximum Entropy Approach to Natural Language Processing," 1996.

[12] J. Pennington,, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *The 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014.

[13] B. Liu, M. Hu and J. Cheng, "Opinion Observer: Analyzing and Comparing Opinions," in *International Conference on World Wide Web (WWW-2005)*, 2005.

[14] D. M. Blei, A. Y. Ng and M. I. Jordan, "Latent dirichlet allocation," *The Journal of machine Learning research,* p. 993–1022, 2003.

[15] A. Purwarianti, A. Andhika, A. F. Wicaksono, I. Afif and F. Ferdian, "InaNLP: Indonesia Natural Language Processing Toolkit," in *ICAICTA*, 2016.

[16] R. Rehurek and P. Sojka, "Software Framework for Topic Modelling with Large Corpora," in *THE LREC 2010 WORKSHOP ON NEW CHALLENGES FOR NLP FRAMEWORKS*, 2010.

[17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research,* vol. 12, pp. 2825-2830, 2011.

[18] N. Okazaki, "CRFSuite: A fast implementation of Conditional Random Fields (CRFs)," 2007. [Online]. Available: http://www.chokkan.org/software/crfsuite/.

[19] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, R. Jozefowicz, Y. Jia, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, M. Schuster, R. Monga, S. Moore, D. Murray, C. Olah, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu and X. Zheng, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," 2015. [Online]. Available: http://tensorflow.org/.

[20] G. Ganu, N. Elhadad and A. Marian, "Beyond the Stars: Improving Rating Predictions using Review Text Content," in *WebDB*.

[21] J. Neville, "Design and Analysis of Machine Learning Experiments," [Online]. Available: https://www.cs.purdue.edu/homes/neville/courses/573/readings/08_design-and-analysis-expts.pdf.