Aspect Based Sentiment Analysis for Review Rating Prediction

Susanti Gojali
School of Electrical Engineering and Informatics
Bandung Institute of Technology
Bandung, Indonesia
susantigojali@hotmail.com

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

***Abstract*—Aspect based sentiment analysis consists of aspect and sentiment extraction, and determination of the sentiment's orientation. In this paper, we propose a system to extract the aspect sentiment pair and compute the rating for each grouped aspect. Our approach starts with selecting the subjective sentences in the reviews. Then, it extracts aspects and opinions from the sentences, and determine the orientation of the sentiment. Aspects are clustered together by using WordNet with the prior knowledge of the aspect categories. Finally, it measures the rating for each aspect category. We evaluate our approach by three criteria: precision, recall, and F1-Measure. The subjectivity classification model and extraction model has F1-Measure value of 0.783 and 0.794 respectively.**

Keywords—aspect-based sentiment analysis; rating prediction; subjectivity classification; aspect and sentiment extraction; aspect aggregation

I. Introduction

Nowadays, online reviews are very common to be found in lots of websites across the internet. One particular review is restaurant reviews. People starting to use online reviews to get a brief explanation of the restaurant or any information they needed. However, the review readers have different kinds of preference about the restaurant. Some would like to find a place delicious meal while others would be more interested on cozy place with good service. On the other hand, all the reviews are written by the reviewers that have different preference from the readers. Therefore, the readers will need to go over all of the reviews to find the essential information.

Some websites such as Trip Advisor [1] and Zomato [2] provide a restaurant rating and reviews. However, the readers will have difficulty finding the relevant information since there are a large number of reviews. Readers will find it easier to know the restaurant aspects with rating based on aspect categories rather than an overall rating on restaurant. However, only few websites provide those rating system. Other than that, the rating is usually provided by the reviewers themselves instead of being extracted from the reviews. Those ratings are subjective and different for each individual. For this reason, the automatic rating generator based on sentiment for each aspect is needed.

In this paper, we propose a system to generate the rating based on Indonesian reviews. Given the reviews of the restaurant, the system will first decide whether the sentence is subjective or objective. Then, the system extracts the potential aspect and sentiment, and determines the orientation of the sentiment. These aspects are grouped to same category, and the system will measure the rating for each category.

In our work, we used supervised learning for subjectivity classification. Three algorithms which are naïve bayes classifier [3, 4], C4.5 [5], and support vector machine are used to classify the sentences. For the information extraction, we will use a method based on sequential learning. Conditional Random Field, one of the supervised approach achieved a good performance [6]. Then, we will group the aspect by utilizing WordNet and give ratings for each category.

The rest of this paper is organized as follows. Section 2 discusses related work. We describe our approach system in Section 3. Section 4 presents our experiment result. Evaluation result and analysis follow in Section 5 and we draw our conclusions in Section 6.

II. Related Work

Subjectivity classification is a process to classify whether a sentence is subjective or objective [7]. There are some technique to classify subjectivity: similarity approach [4] and text classifier [3, 4] . Similarity approach hypothesize that opinion sentence has bigger similarity to opinion sentence than to fact. Yu & Hatzivassiloglou [4] use SIMFINDER to measure the similarity between sentence based on phrase, shared words, and WordNet synset. Text classifier approach use supervised learning, e.g. Naïve Bayes Classifier (NBC) [3, 4]. Wiebe, Bruce, & O'Hara [3] use some feature to train NBC. They use binary feature of adjective, adverb, pronoun, modal, number, punctuation, and the position of the sentence. In our work, we apply Wiebe, Bruce, & O'Hara [3] work. Yu & Hatzivassiloglou [4] also use NBC and multiple NBC to classify opinions and facts. The features include words, bigrams, trigrams, part-of-speech, and counts of positive and negative word in the sentence.

Sentiment analysis is divided into three categories: document level, sentence level, and aspect level. Sentiment analysis which is done at document and sentence level do not represent what the reviewers really like or not. It only represents the whole reviewer's opinion. In this work, we will focus on sentiment analysis on aspect level.

978-1-5090-1636-5/16/$31.00 ©2016 IEEE

Many research have been done on aspect level. There has been a work to extract the aspect by capturing the high frequency keywords [8]. They also extract the infrequent aspect by utilizing the relation between opinion and aspect. Popescu & Etzioni [9] improve the algorithm by removing the noun phrase that may not be the aspect. They compute the pointwise mutual information (PMI) between phrase and the meronymy discriminator of the entity. They improve the precision by 22% while the recall 3% lower.

Another approach is done by using supervised learning. The common method is based on sequential learning, e.g. conditional random field (CRF) and hidden markov model (HMM). Since it is supervised, they need to annotate the aspect and nonaspect manually. Lexicalized HMM is applied to extract aspects and opinions in the sentence [10]. Jakob and Gurevych [11] use CRF and trained it from four different domains. The feature they used are token, POS, short dependency path, word distance, and opinion sentence. Qi and Chen [6] also use linear-chain CRF for opinion mining. They used lexical and POS tag to train the CRF.

Aspects that are extracted can be in large amounts. People tend to use different words to describe same aspect. Aspects that explain the same thing could be grouped in the same category. One study accomplished this task using similarity matrix [12]. They define user-define taxonomy of feature (UDF). They used unsupervised method that Hu & Liu [8] present to get the extracted feature, referred as crude features (CF). They map the crude features to the UDF by measuring the similarity by using WordNet. WordNet was also used to group feature with similar meaning [13]. WordNet is employed to group the feature and two top frequent sense of the word are chosen to find the synonym. Backen et al. [14] also used WordNet based similarity metric called Jcn to cluster the extracted opinion pairs. Zhai et al. [15] also calculate the similarity between two words and tried some similarity calculation algorithm. They conclude Jcn performed best for their task.

For rating calculation, the score can be measured by counting the positive and negative sentence. Ganu et al. [16] use some formula to calculate the rating in scale 1-5. The score can also be measured by using SentiWordNet [14]. SentiWordNet give three scores for each synset: positive sentiment, negative sentiment, and neutrality score. The score is gained by subtracting the negative score from the positive score. The score is aggregated from all synset based on the popularity of the synset.

III. The Proposed Technique

Our goal in this work is to generate rating for each category aspect from the reviews. In this work, the challenge are limited resource and the use of informal language in the reviews. For that reason, we will propose a system using the available resources. There are six main tasks in this system. The architecture is shown in Fig. 1.

The crawler will be used to crawl Trip Advisor's restaurant reviews. We will use Jaunt API to do the crawling. In this work, we only use Indonesian reviews. The example of the review is shown in Fig. 2. Preprocess will do four tasks which are sentence splitting, foreign word translation, sentence formalization, and stop words removal respectively. We use InaNLP tool to do the process.

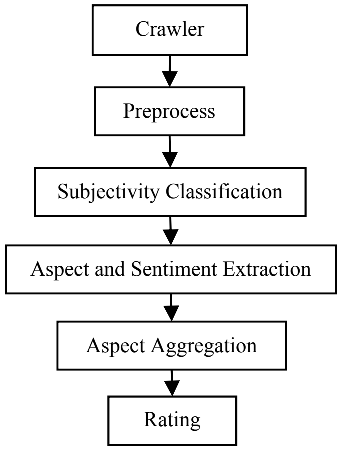

Fig. 1. The Architecture of System

```
(1)Warung Pasta. (2)Cafe yang menyajikan menu aneka olahan pasta yang ber lokasi di Jalan Ganesha, Bandung, ini pastinya udah banyak orang yang tahu. (3)Bisa dibilang cafe ini sudah cukup lama berdiri. (4)Karena berada di samping kampus ITB, tidak heran kebanyakan yang datang adalah kelompok mahasiswa. (5)Tidak hanya pasta yang disajikan di sini, ada juga pizza, nasi keju, sup dan aneka camilan seperti calamaru dan lainnya. (6)Ada juga menu yang bisa dibilang cukup jarang ditemukan di Bandung, yaitu Calzone. (7)Ini semacam sandwich atau bentuknya mirip-mirip pastel kalo di Indonesia, hanya saja ukurannya lebih besar. (8)dessert ckp murah deh.
```

Fig. 2. Review's Example

Sentence splitting process splits review in Fig. 2 into eight sentences. We will use the eighth sentence (`dessert ckp murah deh`) to illustrate the process of foreign word translation, sentence formalization, and stop word removal. Foreign word translation process translate any foreign word in the sentence into Indonesian. We will use our manually construct dictionary to translate the words. In this case, "`dessert`" will be translated to "`makanan penutup`". Sentence formalization transforms the word to formal word. Word "`ckp`" will be formalized to "`cukup`" because it is not formal word. Stop word removal deletes any stop words occur in the sentence based on the dictionary. The detail of the process is shown in Fig. 3.

Foreign word Translation: `makanan penutup ckp murah deh.`
Sentence Formalization: `makanan penutup cukup murah deh.`
Stop word Removal: `makanan penutup cukup murah.`

Fig. 3. Foreign word translation, sentence formalization, and stop words removal example

Reviewers usually use many sentences to write a review. However, not all sentences are subjective sentence. We need to classify each sentence whether it is subjective or objective. The objective sentences will not be used as it does not have opinions. For example, in the Fig. 2, out of eight sentences, only the eight one give an opinion to the restaurant.

We will do supervised learning to classify this sentence. There are seven features that Wiebe, Bruce, & O'Hara [3] used to represent the sentence. The features are binary value of sentence position, pronoun, adjective, number, modal, adverb, and punctuation. In this work, we only used six features, the punctuation feature is not include. For sentence position feature, the feature will have value 1 if the sentence is in the beginning of the paragraph, others will be 0. TABLE I. shows the feature for sentence "makanan penutup cukup murah". Three algorithms will be used to conduct the experiment which are NBC, J48, and Support Vector Machine (SVM). NBC is used because of the good result showed in Wiebe, Bruce, & O'Hara work. SVM and J48 are popular algorithm that used in many domain, so we will use it to compare NBC algorithm. We use Weka [17] implementation to perform NBC, J48 (confidence factor 0.25 and minimum number of object is 2), and SVM (RBF kernel, cost 1.0, ε 0.001).

TABLE I. FEATURE OF SUBJECTIVITY CLASIFICATION EXAMPLE

Feature	Value
Pronoun	0
Adjective	1
Number	0
Modal	0
Adverb	0
Sentence position	0

All sentences that are classified as subjective will have its aspects and opinions words extracted. We define three entities: aspect, positive opinion and negative opinion. TABLE II. shows the example for each entity. We use BIO notation to label the entity. For each entity E, two labels are produced: B-T (beginning token of the entity) and I-T (inside token of the entity). A beginning of entity labeled as B-T while the inside of entity labeled as I-T. In addition, for token outside the entity labeled as O (Other) [18]. For example, the label for the sentence:

```
Steak dan sosis enak sekali

Steak and sausage are very delicious
```

Is

```
Steak<ASPECT-B> dan<Other> sosis<ASPECT-B> enak<O-P_POS-B> sekali<OP_POS-I>

Steak<ASPECT-B> and<Other> sausage<ASPECT-B> are<Other> very<OP_POS-B> delicious<OP_POS-I>
```

CRF algorithm is used to get the best sequence of the tag for a given sentence. We will use two features to train the CRF: the lexical part and its part-of-speech (POS) tags done by InaNLP tool. Mallet tool [19] is used to train and generate the best tag sequence.

TABLE II. ENTITIES AND ITS EXAMPLE

TAG	Description	Example
ASPECT	Aspect	Makanan (*food*), minuman (*beverage*), pelayan (*waiter*)
OP_POS	Positive opinion	Murah (*cheap*), enak (*delicious*), cepat (*fast*)
OP_NEG	Negative opinion	Lama (*slow*), mahal (*expensive*), kotor (*dirty*)

If we have hybrid tags, we could extract the aspect and its matching opinion. For each aspect and its matching opinion, we will determine the real orientation of the sentiment. The sentiment orientation from entity tag does not necessarily become the actual orientation of the sentiment. The presence of negation words could change the orientation of the sentiment. We use the algorithm in [10] and make some modifications. The algorithm to determine the sentiment orientation is presented in Fig. 4.

Algorithm 1 DETERMINE_OPINION _ORIENTATION

```
o = opinion word
initial_opinion_orientation = opinion's orientation
done = FALSE
distance = 1
WHILE (distance <= 5 AND !done ) DO
    IF (o's position – distance) is coordinating conjunction
        done = TRUE
    END IF
    IF (o's position – distance) is aspect
        done = TRUE
    END IF
    IF (o's position – distance) is opinion
        done = TRUE
    END IF
    IF (o's position – distance) is negation word
        done = TRUE
        orientation = opposite (initial_opinion_orientation)
    END IF
    distance++
END WHILE
```

Fig. 4. Algorithm to Determine Opinion Orientation

The algorithm will check the presence of the negative word (e.g. tidak (don't), bukan (not)) within five word distance in front of the opinion word and change the opinion's orientation, except

1. A negation word appear in front of the coordinating conjunction (e.g. dan (and), atau (but))

2. A negation word appear in front of another aspect or sentiment

We use sentence "`steak lezat tapi pastanya tidak enak`" to illustrate the process of opinion orientation determination. In this sentence, we will extract two pairs of aspect sentiment: `<steak, lezat>` and `<pastanya, enak>`. For each pair, we will run the algorithm 1. In the first pair `<steak, lezat>`, the negation word does not found in front of the opinion word. The initial orientation for this opinion is positive and remain the same as there is not any negation word. For the second pair `<pastanya, enak>`, we found a negation word "`tidak`" while we run the algorithm. The initial orientation that is positive is now changed to negative.

Reviewers sometimes use different words to describe the aspect. For example, we can use *movie* and *picture* to describe a film. In our approach, we will aggregate the aspect by using WordNet. However, WordNet can only work in English while the aspect is in Indonesian. First, the aspect will be translated to English using KBBI dictionary. Then, the seeds will be constructed for each category. We define four categories in this work. The categories and the example of the seeds are shown in TABLE III. WordNet is utilized for measuring the similarity of the translated aspect and the seeds. The highest similarity state the category of the aspect. If the similarity is 0 for all categories, the aspect and sentiment pair will classified as other.

TABLE III. ASCPECT CATEGORIES AND ITS SEED

Aspect Category	Example Seeds
Food	Food, beverage, dessert, meal, taste
Service	Service, waiter, waitress
Price	Price
Place	place, atmosphere, ambience

After we have all aspect categories and its aspect, we will calculate the rating for each aspect categories. The rating calculation will follow the equation 1 [16].

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

Variable P/N is the total of positive/negative opinion in the aspect category. The rating is scaled in 1 – 5.

IV. EXPERIMENT

Reviews from Trip Advisor consist of title, text, rating, and reviewer's profile. For this work, only the text part of the reviews were used. We conduct two experiments which are subjectivity classification and entity extraction.

365 reviews from 15 restaurants at Trip Advisor were used for experiment on subjectivity classification. There are 410 objective sentences and 1286 subjective sentences from all reviews. Since the dataset are imbalanced, SMOTE are used. We will perform three algorithms: NBC, J48, and SVM to get the best accuracy by using 10-fold cross validation.

The second experiment will use 605 subjective sentences to perform three scenarios. The detail of data can be seen in TABLE IV. The first scenario only uses lexical feature, the next one uses POS tag feature, and the last one uses both of them. InaNLP tool will be used to produce part-of-speech (POS) for each word in the sentence. Precision, recall, F1-Measure will be measured with 10-fold cross validation to evaluate the model.

TABLE IV. **TOKEN DISTRIBUTION**

Label	Total Token	Example
ASPECT-B	846	interior (*interior*), harga (*price*)
ASPECT-I	256	ruangan (*room*), makanan (*food*)
OP_POS-B	819	banyak (*many*), sangat (*very*)
OP_POS-I	263	pilihan (*choice*), indah (*beautiful*)
OP_NEG-B	152	agak (*quite*), sangat (*very*)
OP_NEG-I	74	mahal (*expensive*), bau (*smelly*)
OTHER	4850	dan (*and*), tapi (*but*)
Total	7257	

TABLE V. shows the result of the experiments. The use of SMOTE does not improve the accuracy of all those algorithms. The three algorithms give nearly the same accuracy, with NBC model being the highest.

TABLE V. EXPERIMENTAL RESULT ON SUBJECTIVITY CLASSIFICATION

Filter	NBC	J48	SVM
No Filter	**78.26%**	**77.67%**	**78.02%**
SMOTE	69.78%	71.23%	70.02%

The results of token classification with CRF algorithm are shown in TABLE VI. The combination of lexical and POS tag feature performs better than using lexical or POS tag alone, with the overall F1-Measure 79.38 %.

From the experiment's results, NBC model will be used for subjectivity classification, and lexical and POS tag will be used as feature for CRF algorithm.

V. EVALUATION AND DISCUSSION

We will use another data to test our model. The data test consists of 100 review (383 subjective and 62 objective). The examples of data test and the labels are shown in TABLE VII. This data test will be used for subjective classification.

The sentence that is classified as subjective from the previous step will be used for token classification as test data. Then, we will extract the aspects and its matching opinions. The results from the system will be compared with our manually constructed labels for aspect and sentiment pair extraction, and aspect aggregation. We will evaluate our approach based on precision, recall, and F1-Measure.

TABLE VI. EXPERIMENTAL RESULT ON ENTITY CLASSIFICATION (10-FOLD-CROSS VALIDATION)

		Lexical	POS tag	Lexical + POS tag
OTHER	P	0.906	0.828	**0.9232**
	R	**0.9594**	0.8963	0.9522
	F	0.9319	0.8608	**0.9375**
ASPECT-B	P	**0.8413**	0.6284	0.8386
	R	0.7518	0.5697	**0.7861**
	F	0.794	0.5976	**0.8115**
ASPECT-I	P	**0.7989**	0.4915	0.7981
	R	0.5898	0.2266	**0.6641**
	F	0.6787	0.3102	**0.7249**
OP_POS-B	P	**0.889**	0.6226	0.8636
	R	0.8217	0.5922	**0.8435**
	F	**0.8541**	0.607	0.8534
OP_POS-I	P	**0.8824**	0.5016	0.8674
	R	0.8555	0.597	**0.874**
	F	0.8687	0.5451	**0.8707**
OP_NEG-B	P	**0.7982**	0.4	0.7667
	R	0.5987	0.0526	**0.6013**
	F	**0.6842**	0.093	0.674
OP_NEG-I	P	**0.8302**	0.3077	0.7143
	R	0.5946	0.0541	**0.6579**
	F	**0.6929**	0.092	0.6849
All Entities	P	**0.8494**	0.5399	0.8246
	R	0.7388	0.4269	**0.7684**
	F	0.7864	0.4436	**0.7938**

TABLE VII. EXAMPLE OF DATATEST

No	Sentence	Label
1	baik secara rasa dan tempat yang cozy untuk bersantai *good flavor and cozy place to relax*	Subjective
2	cafe yang berada di tengah kota *cafe in the city center*	Objective

A. Subjectivity Classification

The result shows that the classifier achieve a high recall and precision on subjective sentences, while on the objective sentences, the classifier performs poorly (TABLE VIII.). One of the sentence that is classified correctly as subjective is "pelayannya ramah dan sigap (*waiters are friendly and spry*)". The sentence has two adjectives which are "ramah (*friendly*)" and "sigap (*spry*)" that express the opinion to "pelayan (*waiters*) ".

One of the subjective sentences that is classified as objective is "semua makanan menggugah selera (*all food are appetizing*)". This sentence does not have an adjective that usually defines the opinion in sentence. Therefore, the classifier classified it as objective sentence. The sentence uses "mengugah selera (*appetizing*)" which are verb and noun to implicitly express the opinion.

Another example of misclassification is "minuman dihidangkan dalam gelas besar. (*Drinks are served in large glasses*)". This sentence is objective but classified as subjective sentence. This is because it has the adjective word "besar (*large*)" so the classifier made the wrong prediction.

TABLE VIII. EVALUATION RESULT ON SUBJECTIVITY CLASSIFICATION

	Subjective	Objective
Precision	83,67 %	59.57 %
Recall	94,6 %	30.11 %

B. Aspect and Sentiment Extraction

For token classification, the accuracy of the label is quite high with 88.48% (TABLE IX. . The results also show that F1-Measure for OP_NEG_I is quite low while for OTHER is high. Many misclassifications occurred and the tokens are mostly classified as OTHER class. This is because of the use of infrequent words to describe the aspects and the opinions. Those infrequent words are then classified as OTHER.

TABLE IX. EVALUATION RESULT ON TOKEN CLASIFICATION

Label	Precision	Recall	F1
ASPECT-B	0.7104	0.7455	0.7275
ASPECT-I	0.4929	0.5475	0.5188
OP_POS-B	0.7524	0.8505	0.7985
OP_POS-I	0.6885	0.8235	0.75
OP_NEG-B	0.6923	0.5373	0.605
OP_NEG-I	0.5926	0.4444	0.5079
OTHER	0.943	0.9243	0.9336
Accuracy			0.8848

One of the example of class ASPECT classified as OTHER is in the sentence "pernah order gurame bakar, guramenya sedikit overburn (*order grilled carp, the crap slightly over burn*)". In that sentence, the aspect is "guramenya (*the crap*)" however the model cannot detect the aspect. It is caused by that word never appeared in training data. The opinion in that sentence is "sedikit overburn (*slightly over burn*)". Those words also never appear in the data and also classified as OTHER.

In contrast, all tokens in the sentence "tempatnya enak, suasananya juga nyaman, menu makanan bervariasi (*the place is nice, the atmosphere is also comfortable, and a lot of variation in the menu*)" are correctly classified. The reason behind this is all of the tokens have occurred in the training data and the model can easily recognize the pattern in the sentence.

For extraction aspect and sentiment pair, the system does not show a good result TABLE X. The system performs quite poor mostly because of the misclassifications that occur in the token classification step. One of the example is "rasa makanan tergolong biasa". The phrase "rasa makanan" labeled as aspect and "tergolong biasa" labeled as positive opinion. Then, the system extracted the pair of <rasa makanan, tergolong biasa> while the correct pair is <rasa makanan, biasa>.

TABLE X. EVALUATION RESULT ON ASPECT SENTIMENT PAIR

Precision	Recall	F1
0.667	0.552	0.582

Aggregation aspect step also perform poorly (TABLE XI. The performance decreases just a little bit compared to aspect and sentiment evaluation. An example of miscategorization is an aspect "`udang bakar` (*`roasted prawn`*)" categorize as place.

TABLE XI. EVALUATION RESULT ON ASPECT AGGREGATION

Precision	Recall	F1
0.637	0.525	0.554

After the aspect extraction step, we measured the rating for each category. For example, given the review:

```
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 have a lot to learn.
```

The system will produce the rating for each category based on the review. The output of the system is shown in Fig. 5.

```
Place : 5 star
    (+) pemandangan => bagus ( view => nice)
    (+) tempat => nyaman ( place => comfortable)
Price : 5 star
    (+) harga => terjangkau  (price => affordable)
Food : 1 star
    (-) minuman cocktail => kurang enak (cocktail =>
        was not too good)
```

Fig. 5. Output of the System

VI. CONCLUSIONS

The restaurant rating predictor system can be built in some steps: preprocessing the reviews, selecting the subjective sentences, extracting the pair of aspect and sentiment and determining the orientation of the sentiment, aggregating the aspect into the same group, and measuring the rating. NBC algorithm is the best model for subjectivity classification with F1-Measure value of 0.783. The best feature for CRF model is the combination of lexical and POS tag with F1-Measure value of 0.794.

The performance of the system can be improved by adding new feature to NBC model and CRF model. Adding training data can also improving the accuracy of the model. To improve the performance of extraction process, using some rules can be used to extract the aspect and opinion pair.

REFERENCES

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

[2] "Zomato," [Online]. Available: https://www.zomato.com.

[3] J. M. Wiebe, R. F. Bruce and T. P. O'Hara, "Development and Use of a Gold-Standard Data Set for Subjectivity Classifications," in *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, 1999.

[4] H. Yu and V. Hatzivassiloglou, "Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences," in *Proceedings of the 2003 conference on Empirical methods in natural language processing*, 2003.

[5] J. R. Quinlan, in *C4. 5: programs for machine learning*, Elsevier, 2014.

[6] L. Qi and L. Chen, "A linear-chain CRF-based learning approach for web opinion mining," in *Web Information Systems Engineering*, 2010.

[7] B. Liu, Sentiment analysis and opinion mining, 2012, pp. 1-167.

[8] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004.

[9] A. M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in *Natural Language Processing and Text Mining*, Springer London, 2007, pp. 9-28.

[10] W. Jin and H. H. Ho, "A Novel Lexicalized HMM-based Learning Framework for Web Opinion Mining," in *Proceedings of the 26th Annual International Conference on Machine Learning*, 2009.

[11] N. Jakob and I. Gurevych, "Extracting Opinion Targets in a Single- and Cross-Domain Setting with Conditional Random Fields," in *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 2010.

[12] G. Carenini, R. T. Ng and E. Zwart, "Extracting knowledge from evaluative text," in *Proceedings of the 3rd international conference on Knowledge capture*, 2005.

[13] B. Liu, M. Hu and J. Cheng, "Opinion Observer: Analyzing and Comparing Opinions," in *Proceedings of the 14th international conference on World Wide Web*, 2005.

[14] W. Bancken, D. Alfarone and J. Davis, "Automatically detecting and rating product aspects from textual customer reviews," in *Proceedings of DMNLP Workshop at ECML/PKDD*, 2014.

[15] Z. Zhai, B. Liu, H. Xu and P. Jia, "Clustering Product Features for Opinion Mining," in *Proceedings of the fourth ACM international conference on Web search and data mining*, 2011.

[16] G. Ganu, N. Elhadad and A. Marian, "Beyond the Stars: Improving Rating Predictions using Review Text Content," in *WebDB*, 2009.

[17] H. Mark, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, "The WEKA Data Mining Software: An Update," 2009.

[18] C. C. Aggarwal and C. Zhai, Mining Text Data, Springer Science & Business Media, 2012.

[19] A. K. McCallum, "MALLET: A Machine Learning for Language Toolkit.," 2002. [Online]. Available: http://mallet.cs.umass.edu.