Aspect Evaluation by using Overall Rating and Category Characteristics of Reviews

Hatsuho Sakai#1 and Ryosuke Saga*2

#Graduate School of Engineering, Osaka Prefecture University
1-1, Gakuencho, Naka-ku, Osaka, Japan
1 sakai@mis.cs.osakafu-u.ac.jp
2 saga@cs.osakafu-u.ac.jp

***Abstract*-This paper presents a method for estimating aspect rating in reviews. Aspects are evaluated using evaluative words. The overall rating of reviews is used to estimate the rating of aspects. We assume that reviews with words expressing high evaluation possess high overall rating. We estimate evaluative words for each category because some of these words express different meanings in different categories. We determine the score of an aspect from the rating of evaluative words. The approach is validated by estimating the values of aspects by using reviews collected from kakau.com and comparing them with the original aspect ratings. Results indicate that the proposed approach can estimate aspect rating in certain cases.**

***Keywords*-Opinion Extraction, Review Mining, Sentiment Analysis, Aspect Extraction, Latent Rating Analysis**

Introduction

Currently, users refer to online review sites or comparison shopping websites when they consider buying something. In some comparison sites, overall and aspect ratings are displayed. Aspects are attributes for evaluating products, such as design, graphic, and usage. Fig. 1 shows an example of a review. The product name is written in boldface (Camera A) on the top part, and the product category is written with an underline (Digital Camera). The number of big black star expresses the overall rating. Aspect ratings are expressed with the number of small black stars. Review texts are added to the right side of the space where aspect ratings are displayed. A user can easily compare products by using aspect ratings. In some cases, no aspect ratings are available for users. As such, the user attempts to know the reputation of the product based on review texts. However, determining a desired opinion from the Internet tends to be difficult.

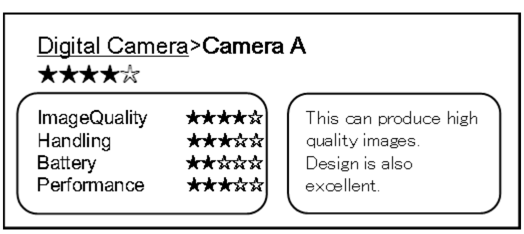

Fig. 1. Example of a review.

Aspect-evaluation studies are proposed from the background. Methods focused on aspect and evaluative words are proposed to extract the reputation of aspects. Evaluative words, such as "good" or "bad," indicate the quality of aspect. A technique using overall rating of reviews is proposed to estimate aspect ratings and evaluative words. However, some words have different usage by category. For example, "A hot new book" seems like a favorable comment, but "CPU becomes too hot" seems to be a negative one. Hence, some evaluative words that express favorable meanings may also be used to express negative sentiments. Such rating cannot be defined with consistency.

In this paper, we propose a method for estimating aspect rating from reviews. We assume that favorable words that appear in reviews indicate high overall ratings. We set the rate of an evaluative word to the overall rating, in which the word appears the most frequently. We also use category characteristics. Category is used in review sites to classify products, such as cameras, smart devices, and games. We assume that some evaluative words express different meanings in different categories. We estimate the value of evaluative words by each category. Finally, we determine the aspect ratings from evaluative words. To evaluate the proposed method, we conduct experiments and evaluate aspects in reviews collected from various websites.

II. RELATED WORK

Methods for estimating the polarity of a document are studied to analyze an opinion written in the document [1, 2]. Multiple opinions may appear in a document. Therefore, approaches in judging polarities from a sentence are proposed [3, 4]. Furthermore, in some documents, one sentence evaluates plural aspects, or plural sentences evaluate an aspect. As such, aspect extraction and evaluation are suggested to estimate the polarity or rating of an aspect. Kobayashi et al. [5] proposed an approach for aspect extraction focused on subject, aspect, and value. Subject is a product or service. Value is expressed through evaluative words. Opinion is deemed to contain three elements, namely, subject, aspect, and value. Consider the example, "The design of product is excellent;" "product" is the subject, "design" is the aspect, and "excellent" expresses the value. Opinions are collected as a triplet <Subject, Aspect, Value> by using co-occurrence patterns. The seed dictionaries of aspect and value are set. When a sentence is applied to a pattern, a candidate aspect is then extracted. If the candidate is not included in the dictionary, then the candidate is added.

Arjun et al. [6] extracted and classified aspects from a sentence through semi-supervised learning. Tao et al. [7] proposed aspect extraction using two semi-supervised topic models.

Samuel et al. [8] suggested an approach that uses evaluative words when polarity was added beforehand to judge the polarity of the aspects. In this approach, unsupervised learning is used to extract sentences that evaluate aspects. Furthermore, the co-occurrence degrees of the same polarity of words are assumed as high. The polarity of evaluative words is judged based on co-occurrence degree with the word that the polarity touched. The polarity of aspects is estimated based on the polarity of the evaluative word.

Fu et al. [9] calculated the similarity of evaluation words by using a word net and performed polarity classification of aspects.

Hongning et al. [10] proposed an aspect rating method by calculating the weight of an evaluative word that appeared in the reviews of the hotels and estimated the ratings of aspects. The approach can be classified into two stages. On the first stage, sentences are identified to evaluate an aspect and an aspect is expressed using plural words. For example, the aspect called "the design" may be expressed through words, such as "form," "size," and "color." A word is extracted to express an aspect by using bootstrap method. Four seed words are set for seven aspects, and bootstrap method is applied to every sentence to identify sentences that evaluate aspects. On the second stage, aspect evaluation is performed. According to Hongning, the overall rating of a review is the total of the weighted value of a rating of aspects. Furthermore, the rating of an aspect is assumed as the sum of the ratings of evaluative words. Based on this hypothesis, the ratings of evaluative words and aspects are estimated.

Jianxing et al. [11] considered the overall rating as the sum of the rating of aspects and performed aspect rating.

Some evaluative words are unsuitable for uniquely defining the rating. In this paper, we propose a method to estimate the rating of aspects by using category characteristics.

III. Aspect Evaluation

In this paper, we determine a score of an aspect from the rating of an evaluative word. Aspect is expressed by words, such as "design," "graphic," and "usage." We assume that some words express an aspect. For example, "looks" and "appearance" may express the same aspect. In this paper, we propose methods wherein aspect word extraction is performed beforehand.

Fig. 2 shows the proposed method. First, we perform dependency structure analysis to clarify the relationships between words and their described aspects. Second, we identify evaluative words. Third, we estimate the rating of evaluative words based on the overall rating and category characteristics. We assume that some evaluative words express different meanings in different categories. Finally, we determine a rating of aspects from the evaluative words. For each step, we can apply several approaches (Fig. 2).

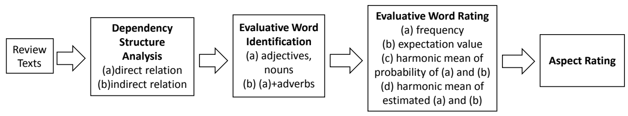

Fig. 2. Process of aspect evaluation.

A.Dependency Structure Analysis

We perform dependency structure analysis [12] to clarify the relationships between words and their described aspects. For example, in the sentence "The design is cool," "cool" modifies "design." We perform the division of the sentence in the following rules for every lines.

1) *Divide the line from the head of the line to the point that a period appeared as one sentence*
2) *Connect the next line when a comma appears in the end of the line*
3) *Divide the entire line as one sentence when a comma does not appear in the end of the line*

We assume that evaluative words directly modifying aspects are useful. However, in some cases, we could not estimate the aspect ratings by using directly modifying words. We propose two approaches to extract evaluative words modifying aspects.

1) *Consider only directly modifying words*
2) *Consider indirectly modifying words*

B. Evaluative Word Identification

We identify the specific parts of speech as evaluative words in review texts. We assume that evaluative words are adjectives, nouns, and adverbs. However, some adverbs that express favorable reputation could not be determined. In some cases, adverbs are unsuitable for estimating ratings. We propose two patterns of the part of speech to identify evaluation word.

1) *Adjectives, Nouns*
2) *Adjectives, Nouns, Adverbs*

C. Evaluative Word Rating

We estimate the ratings of evaluative words by using the overall ratings of reviews. We assume that evaluative words have different meanings in different categories. We estimate the ratings by each category. The overall rating is an integer from 1 to 5. We propose four approaches to estimate the ratings of evaluative words.

1) Appearance Frequency: We assume that in reviews with high overall rating, favorable words appear frequently. In each category, we set the rate of evaluative words to the overall rating, in which the evaluative word appears the most frequently. The score of evaluative word r_a is defined as follows:

$$r_a = arg\ max_{1 \leq x \leq 5} \frac{C_{Ex}}{C_x} \tag{1}$$

χ is the overall rating of reviews. E is an evaluative word. C_{EX} is the number of reviews, in which E appears;

and the overall rating is x. C_X is the number of reviews, in which the overall rating is χ.

2) Expectation Value: We assume that a favorable word appears in high overall rating reviews. In each category, we set the rate of evaluative words to the expectation value of the overall rating, in which the evaluative word appears. The score of E r_{exp} is defined as follows:

$$r_{exp} = \sum_{1 \le x \le 5} x \frac{C_{E,x}}{C_E} \qquad (2)$$

$C_{E,x}$ is the number of occurrences, wherein □ appears in the reviews, in which the overall rating is x. C_E is the number of occurrence of E.

3) Expectation Value of Harmonic Mean of Probability: In some cases, $C_{E,x} / C_E$ calculated with 2) increases as C_X increases. If C_X is extremely high, then $C_{E,x}$ increases even if appearance frequency of is low. In such cases, r_{exp} approximates x. To minimize the influence of the number of reviews, we use the expectation value of harmonic mean of C_{EX}/C_X and $C_{E,x} / C_E$. The score of r_{prb} is defined as follows:

$$r_{prb} = \sum_{1 \le x \le 5} 2x \frac{C_{Ex}}{C_x}\frac{C_{E,x}}{C_E} / \left(\frac{C_{Ex}}{C_x} + \frac{C_{E,x}}{C_E}\right) \qquad (3)$$

4) Harmonic Mean of Estimated Value of 1) and 2): In some cases, r_{exp} becomes near to x which C_X is big. To reduce the influence of the number of reviews, we use harmonic mean of r_a and r_{exp}. The score of r_{est} is defined as follows:

$$r_{est} = 2 r_a r_{exp} / (r_a + r_{exp}) \qquad (4)$$

D. Aspect Rating

We extract evaluative words for evaluating aspect A in review *D*. We calculate the mean value of evaluative words r_{AD} as a score of an aspect. r_{AD} is defined as follows:

$$r_{AD} = \frac{1}{|W|} \sum_{w \in W} r_w \qquad (5)$$

W is the set of evaluative word w , which evaluates A. r_w is the rating of w.

IV. Experiments

A. Data Sets

We use reviews collected from kakaku.com [13]. This website is a Japanese comparison shopping website. We collected 6,021 reviews on July 15, 2015. Reviews are written for 2,041 products and are classified into 30 categories. About 4,599 users wrote the reviews. The number of reviews by each overall rating is shown in Table 1. The number of reviews and average lines of each category are shown in Fig. 3. We use the morphological analyzer by MeCab [14] and dependency structure analyzer by Cabocha [15] for dependency structure analysis.

In section 3, we propose two approaches for dependency structure analysis, two for evaluative word identification, and four for evaluative word rating. We use 16 combined approaches for experiments. We estimate the ratings of aspect in every review and compared it with the ratings that users touched. The contents and number of aspects are different by category. In addition, we use aspect words set by hands.

We use correlation, average, and variance of absolute error of estimated ratings and original ratings to provide the performance index. We calculate these indices by combining approaches and categories. A p value less than 0.05 was considered statistically significant.

TABLE I
THE NUMBER OF REVIEWS BY EACH OVERALL RATING

Overall Rating	The Number of Reviews
1	265
2	238
3	580
4	1891
5	3047

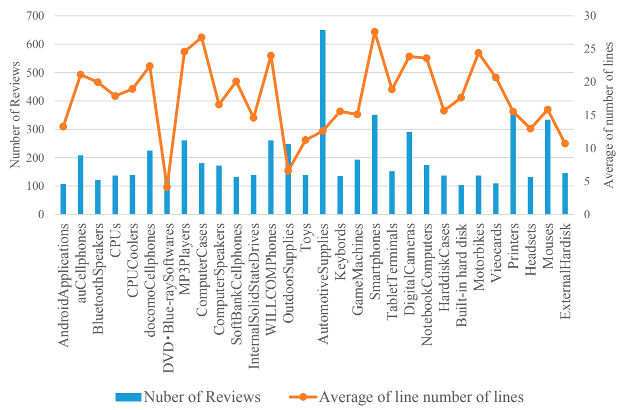

Fig. 3 Number of reviews and average lines of each category.

B. Performance of Aspect Evaluation Focused on Approach

1) Experiment Result of each Approach

Correlations of each approach are shown in Fig. 4. The averages of absolute error are shown in Fig. 5. Variances of absolute error are shown in Fig. 6. The red point in Figs. 4, 5, and 6 shows the mean value. The numbers of evaluative words by overall ratings are shown in Table 2, in which r is the overall rating of evaluative words. A1 shows the performance when we considered only the directly modifying words for dependency structure analysis. A2 considers indirectly modifying words. B1 shows the performance when we used adjectives and nouns as evaluative words. By contrast, B2 uses adjectives, nouns, and adverbs. C1 shows the performance when appearance frequency is used for rating of evaluative words. C2 uses expectation value, C3 uses expectation value of harmonic mean of probability, and C4 uses harmonic mean of the estimated value.

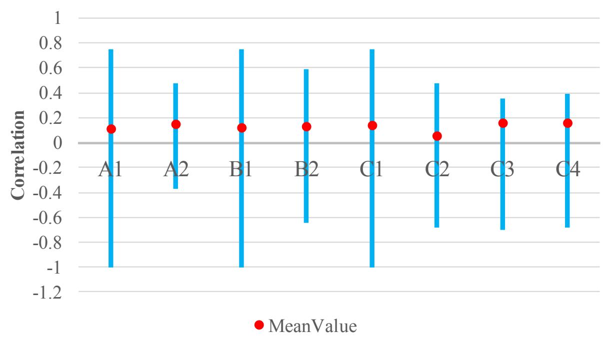

Fig. 4 Correlation of each approach.

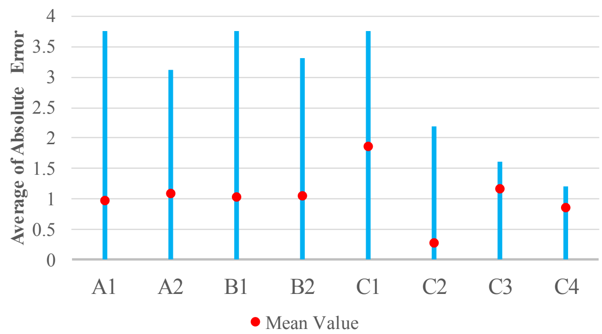

Fig. 5 Average of absolute error of each approach

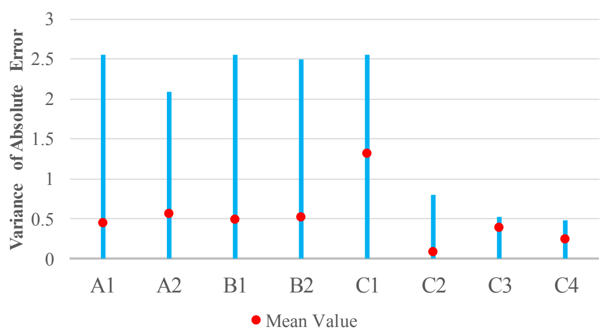

Fig. 6 Variance of absolute error of each approach

TABLE II
THE NUMBER OF EVALUATIVE WORDS BY EACH APPROACHES AND OVERALL RATING

Overall Rating	Approach							
	A1				A2			
	C1	C2	C3	C4	C1	C2	C3	C4
0≦r<1	0	0	9	0	0	0	15	0
1≦r<2	1097	301	760	464	1668	454	1181	688
2≦r<3	942	441	1647	1017	1445	677	2487	1610
3≦r<4	1522	1955	2532	2855	2250	2977	3613	4187
4≦r<5	2331	5366	3115	3727	3530	7826	4638	5449
r=5	2784	2784	2784	2784	4266	4266	4266	4266

A1 shows a higher maximum of the correlation than A2, and the minimum is small. A2 shows a higher mean value of the correlation. A1 exhibits higher absolute error average and the maximum of the absolute error dispersion. A2 presents a high mean value.

The maximum correlation of B1 is higher than that of B2, and the minimum is small. The mean of B2 is high. Absolute error average and the maximum of the absolute error dispersion of B1 are high. The mean of B2 is also high.

C1 shows the maximum of the correlation compared with other approaches, and the minimum is the smallest. The absolute error average and the absolute error dispersion are the highest. In addition, the number of evaluative words that estimated less than 2 and more than 1 is higher than that in other approaches. Few words are rated more than 4.

C2 shows the smallest correlation mean among all other approaches. Many evaluation words are estimated to be more than 4.

C3 and C4 show the smallest maximum of the correlation, but the minimum and mean values are high. Conversely, the average and maximum of the absolute error dispersion are low. In addition, estimated evaluative words are higher than that of other approaches that estimated less than 4 and more than 2

2) Discussion

First, we consider the dependency structure analysis. Evaluative words that modify aspects directly are more useful but present many errors. When an evaluative word modifies a different word, it is effective in revising the rating of the evaluative word by using the rating of a word modifying it.

Next, we consider identifying the evaluative words. Adverbs are deemed useless for evaluative word rating. We assume that adverbs emphasize evaluative words. The precision of estimation improves by using adverbs for rating the revision of the evaluative words.

Then, we consider evaluative word rating. The approach using appearance frequency chooses the highest one in the case that the probability of the plural overall rating is almost the same. The approach using appearance probability, on the other hand, is considered to easily cause errors. When we estimate the rating, dividing reviews into positive (overall rating 1~2) and negative set (overall rating 4~5) and judging the polarity of evaluative words are considered effective. Then, we calculate the rating of evaluative words from the appearance frequency of each overall rating afterwards.

The approach using expectation value shows the evaluative words with highly estimated rating. Based on the reviews we collected, those with minimal influence have low overall ratings. When we use expectation, we need to calculate the rating with the same number of reviews by every overall rating, but reviews may be not collected in some categories. We assume that calculating the rating by using the review of similar category is effective.

The approach using harmonic mean is used to reduce the influence of the number of reviews of every overall rating. However, the maximum correlation is low. As such, the approach is less effective compared with other approaches.

C. Performance of Aspect Evaluation focused on Category

1) Experiment Result of each Category

Correlations of each category are shown in Fig. 7. The printer shows the highest in terms of the maximum value of correlation, which is 0.747. CPUs, hard disk cases, internal hard disks, and headsets show more than 0.5 in maximum. Toys, video cards, and CPU coolers show small value in maximum correlation. The category is lower than −0.5 with toys, video cards, and keyboards in the minimum of the correlation. Au cell-phones, docomo cell-phones, and bluetooth speakers exhibit the highest correlation mean. Toys, video cards, and CPU present a small mean correlation.

The averages of absolute error are shown in Fig. 8. MP3 player, outdoor supplies, and motorbikes show the highest value in the average of absolute error. The category, SSD, WILLCOM Phones, and CPU coolers have a small value. Outdoor supplies, motorbikes, and MP3 players show the highest mean absolute error. Docomo cell-phone, SSD, and WILLCOM Phones have small values.

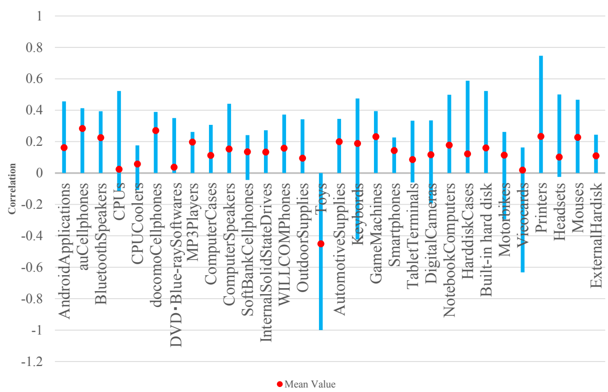

Fig. 7 Correlation of each Category

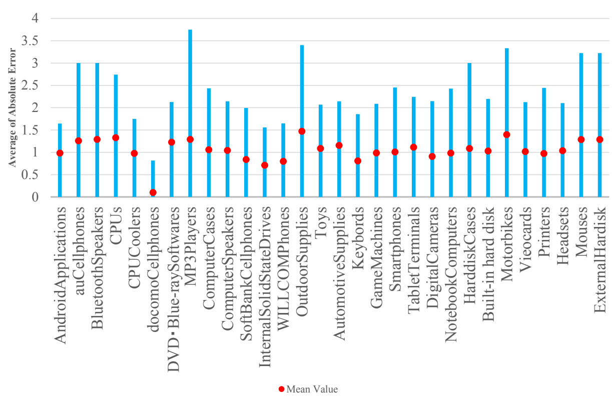

Fig. 8 Average of absolute error of each Category

2) Discussion

The printer has the highest maximum coefficient of correlation. The number of reviews on printers came in second with 372 reviews, as shown in Table 4. The category wherein the maximum coefficient of correlation is considered big (WILLCOM Phones, au cell-phones, and docomo cell-phones) has more than 200 reviews. Categories with such number of reviews can be considered to have a coefficient of correlation with the estimated value and a user set that becomes higher.

Car supplies have a large number of reviews, but its coefficient of correlation is not so high. The range of products included in the car supplies is wide. For example, car navigation systems, car seats, and audio systems are included. Therefore, the evaluation that the evaluative word expressed is not constant. Thus, we cannot attach a proper evaluation.

The toy has low correlation by all approaches. The average number of line of reviews with toys is 6.581, which is the second lowest value. In addition, DVD and Blu-ray soft wares have the fourth smallest mean correlation and contains the least number of the average of linage. Few evaluative words in the review have few lines, and the precision of the aspect rating is considered low. In addition, the range of products included in the toys is also wide and includes figure dolls, education toys, and board games. Therefore, the correlation is extremely low.

V. Conclusions

In this paper, we propose a method for evaluating aspects in reviews by using overall rating and category characteristics. Aspects are attributes that evaluate products and are described by evaluative words. Methods for aspects and evaluative words rating by using overall rating of reviews are proposed. However, some evaluative words have different meanings in different categories. Such a rating of a word could not be defined with consistency. To solve this problem, we estimated the rating of evaluative words by each category. To evaluate this method, we conducted experiments. We estimated the ratings of aspects in the reviews collected from kakaku.com and compared them with the original aspect ratings.

Evaluative words that modify aspects directly are more useful. Adverbs modifying adjectives and nouns are not useful for evaluative word rating. The approach using appearance frequency had the highest in the correlation, but many errors also occurred. The approach using expectation value showed the smallest mean of correlation than in other approaches. The approach using harmonic mean showed the smallest value in maximum correlation, but the value of the minimum and the mean were large.

When we focus on categories, the correlation is high in the category with much number of reviews, and a coefficient of correlation was low as for the linage few categories of the review upon accumulation. In addition, a coefficient of correlation lowers when the width of the product included in the category is wide.

As a future task, we need to revise the rating of evaluative words by using adverbs. We assumed that adverbs are used to emphasizing evaluative words. In addition, in this paper, we did not perform aspect extraction. We need to extract aspects automatically.

REFERENCES

[1] D. T. Peter and National Research Council of Canada, Ottawa, Ontario, Canada, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, Philadelphia, Pennsylvania, 2002, pp. 417-424.

[2] T. Maite, B. Julian, T. Milan, V. Kimberly, and S. Manfred, "Lexicon-based methods for sentiment analysis," *MIT Press Cambridge*, vol. 37, no. 2, pp. 267–307, 2011.

[3] Y. Hong and H. Vasileios, "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 Processin*g, 2003, pp. 129 - 136.

[4] S. M. Kim and E. Hovy, "Analyzing Predictive Opinions on the Web," in *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, Prague, Czech Republic, 2007, pp. 1056 - 1064.

[5] K. Nozomi, M. Yuji, T. Kenji, and F. Toshikazu, "Collecting Evaluative Expressions for Opinion Extraction," in *Proceedings of the First International Joint Conference on Natural Language Processing*, Hainan Island, China, 2005, pp. 596 - 605.

[6] M. Arjun and L. Bing, "Aspect Extraction Through Semi-supervised Modeling," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1*, Jeju Island, Korea, 2012, pp. 339 - 348.

[7] W. Tao, C. Yi, L. Ho fung, Y. K. L. Raymond, L. Qing, and M. Huaqing, "Product aspect extraction supervised with online domain knowledge," *Knowledge-Based Systems*, vol. 71, pp. 86-100, 2014.

[8] B. Samuel and E. Noemie, "An Unsupervised Aspect-sentiment Model for Online Reviews," in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Los Angeles, California, 2010, pp. 804 - 812.

[9] F. Xianghua, L. Guo, G. Yanyan, and W. Zhiqiang, "Multi-aspect Sentiment Analysis for Chinese Online Social Reviews Based on Topic Modeling and HowNet Lexicon," *Elsevier Science Publishers B. V*, vol. 37, pp. 186 - 195, 2013.

[10] W. Hongning, L. Yue, and Z. Chengxiang, "Latent Aspect Rating Analysis on Review Text Data: A Rating Regression Approach," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, DC, USA, 2010, pp. 783 - 792.

[11] Y. Jianxing, Z. Zheng Jun, W. Meng, and C. Tat Seng, "Aspect Ranking: Identifying Important Product Aspects from Online Consumer Reviews," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, Portland, Oregon, 2011, pp. 1496 - 1505.

[12] K. Taku and M. Yuji, "Japanese Dependency Structure Analysis Based on Support Vector Machines," in *Proceedings of the 2000 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora: Held in Conjunction with the 38th Annual Meeting of the Association for Computational Linguistics - Volume 13*, Hong Kong, 2000, pp. 18 - 25.

[13] "Kakaku." [Online]. Available:http://kakaku.com.

[14] MeCab, "Yet Another Part-of-Speech and Morphological." [Online]. Available:http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html.

[15] CaboCha, "Yet Another Japanese Dependency Structure Analyzer." [Online]. Available:https://taku910.github.io/cabocha.