On-site Impression Grasping System Using SNS Location Information and Sentiment Analysis

Ryosuke Yamano
AAII, Department of Data Science
Musashino University
Tokyo, Japan
s2022035@stu.musahino-u.ac.jp

Thatsanee Charoenporn
AAII, Department of Data Science
Musashino University
Tokyo, Japan
thatsane@musashino-u.ac.jp

Virach Sornlertlamvanich
AAII, Department of Data Science
Musashino University
Tokyo, Japan
Faculty of Engineering
Thammsat University
Pathum Thani, Thailand
ORCID: 0000-0002-6918-8713

Abstract—In the recent years, information from such as Social Networking Service (SNS) is overflowing. It draws a great attention from research community in efficiently collecting and analyzing what is being shared in real time. Extracting topics in SNS and analyzing the emotional expression related to those topics are one of the means to know the trends of social interest. In the conventional sentiment analysis, the sentiment of a sentence is estimated on a word-by-word basis, which tends to give undesired results. In many cases, such as negative auxiliary verbs, adverbs, and adjectives related to emotional words are ignored in configurating the emotional expressions. In this research, a method of natural language processing (NLP) in sentiment analysis is performed phrase by phrase by combining the results of active text analysis with the location information. We propose a method to grasp the social impression on an event by improving the phrase level sentiment analysis combining to the location information.

Keywords— Social Networking Service (SNS), sentiment analysis, social interest, emotional expression, phrase structure

I. INTRODUCTION

We use the word emotion polarity correspondence table to identify the impression of the analysed text data from Twitter. The problem in word by word emotional analysis can be solved by parsing the input text and defining the scope of the modification of emotional word. Most of the available sentiment analysis in standard library use the pre-defined sentiment word lists based on SentiWordNet [5] and alike. The advantage of using SentiWordNet is because it is created based on WordNet where the word semantic relations are useful for further computation, for instance, synonym, antonym, hypernym, hyponym, meronym, holonym and entailment. The semantic relations can help in grouping or looking for alternative expression. There are three types of sentiment score defined in SentiWordNet, those are objective, positive, and negative. A word has value 1.0 for the objectivity scores indicates that the term is objective, while 0.0 means that the term conveys some strong sentimental (positive or negative) meaning.

TextBlob [1] is a lexicon-based approach for sentiment analysis. The python library TextBlob actively uses Natural Language ToolKit (NLTK) to achieve its tasks. The sentiment is defined by its semantic orientation and the intensity of each word in the sentence. This requires a pre-defined dictionary classifying negative and positive words. Generally, a text message will be represented by bag of words. After assigning individual scores to all the words, final sentiment is calculated by some pooling operation like taking an average of all the sentiments[2].

In this research, we introduce PN value of the Word Emotion Polarity Correspondence Table to assign the word polarity for computing the phrase level sentiment analysis. The PN value of the Word Emotion Polarity Correspondence Table is a positive-negative value shows whether each word has a positive (example: beautiful, excellent) or negative (example: dirty, inferior) meaning. [2] We need a syntactic parser to indicate the word part-of-speech and the phrase structure for extracting the scope of modification. The content words in the tweets are assigned by their sentiment values. The scope of modification within a phrase is used to disambiguate the modification of the sentiment values to their head words. Finally, the sentiment of the tweets is summed up to indicate the topic of interest and when combined with the location information of the tweets, the location based sentiment analysis of SNS can be indicated.

II. LITERATURE REVIEW

It is reported that the information such as SNS is overflowing and shared in real time, which is efficiently collected and analyzed. Most of the available sentiment analyses in the standard library are performed a lexical-based manner. Sentiment Classification for Hotel Booking Review Based on Sentence Dependency Structure and Sub-Opinion Analysis [6] parses and focuses on dependent words.

In this study, the PN value of the Word Emotion Polarity Correspondence Table is used in the sentiment analysis to grasp the sentiment value of the main words in each phrase of the sentence. The proposed sentiment analysis approach is a kind of phrase-based method as described in the steps below.

- Predicates (adjectives and verbs) of the phrase
- Presence or absence of adverbs of the predicates
- Presence or absence of the negator

The above three steps lead to phrase-based sentiment analysis of sentences and will be treated as phrases hereafter. For the calculation method when there is an adverb modifying the predicate or there is a negator, we apply the TextBlob sentiment calculating polarity and subjectivity methodology [1].

[1]https://textblob.readthedocs.io/en/latest/advanced_usage.html#sentiment-analyzers

[2] https://towardsdatascience.com/my-absolute-go-to-for-sentiment-analysis-textblob-3ac3a11d524

The PN value of the Word Emotion Polarity Correspondence Table shows whether each word has a positive (example: beautiful, excellent) or negative (example: dirty, inferior) meaning. The polarity determination is calculated automatically using the vocabulary network. Words assigned values in the range -1 to +1 and assigned values close to -1 are assumed to be negative, and values close to +1 are assumed to be positive. The vocabulary is extracted from the Iwanami dictionary.

Tohoku University's Japanese Evaluation Polarity Dictionary [4] is created using a part of the Word Emotion Polarity Correspondence Table. The Japanese Evaluation Polar Dictionary [4] assigns subjectivity and objectivity to emotional words. A list of about 5,000 evaluation expressions (evaluation value expression dictionary [2]) collected mainly from words that can be predicates has been partially reorganized, and evaluation polarity information has been added manually. The evaluation polarity tags are as follows. There are four categories.

- Positive / subjective
- Positive / objective
- Negative / subjective
- Negative / objective

However, we use Word Emotion Polarity Correspondence Table for sentiment analysis because we do not parse the sentence with context and do not take subjectivity and objectivity in concern.

III. Methodology

The data handled in this study are the Twitter text data, location information, and Word Emotion Polarity Correspondence Table. There are 55,125 words in the word emotion polarity correspondence table. Each word is labeled in PN (Positive / Negative) value in the range of -1 to 1. The numbers of positive and negative words are unbalance. There are 5,122 positive words and 49,983 negative words. Most of the words are labeled in negative value. Therefore, the result of sentiment analysis has a trend to return negative a value since the unregistered words are ignored and the phrase without a sentiment word is dropped out.

Regarding the positioning of the subject and the predicate, the words that are related to nsubj are extracted through the dependency tag of spacy. Among them, the dependency source is the predicate and the dependency destination is the subject.

The tweet text data are collected based on the location information of the place where is interested in. From the collected tweets, the subject and predicate are extracted to study the sentiment expression. The collected text data is analyzed one sentence each. The tweets without subject or predicate are excluded in this step. As a result, the target text data containing not only the subject and the predicate but also the adverb modifying the predicate and the negation word are extracted for further analysis.

The PN values of the subject and predicate are output by looking up in the Word Emotion Polarity Correspondence Table. Since there are only 55,125 words in the Word Emotion Polarity Correspondence Table, 0 is given to the word that does not exist in the database. For the calculation of the PN value of the subject and the predicate, the PN value of the phrase is calculated by averaging the PN value of the subject and the predicate, as shown in (1).

$$((\textit{PN value of subject} + \textit{PN value of predicate})) / 2 \quad (1)$$

When an adverb is extracted or a negative sentence is judged, the calculation method of TextBlob Sentiment: Calculating Polarity and Subjectivity is applied. In case of adverb, the PN value of the phrase is the sum of PN value of adverb and the modified predicate, as shown in (2)

$$(\textit{PN value of adverb} + \textit{PN value of predicate}) \quad (2)$$

In case of negation phrase, the PN value of the phrase is the multiplication of total phrase PN value and -0.5, as shown in (3).

$$(\textit{PN value of the phrase} \times (-0.5)) \quad (3)$$

The total PN value of the sentence is the sum of the PN value of the phrasal components of the sentence.

Regarding the calculation method, we considered treating the PN value of the extracted adverb as an absolute value, normalizing it to 1.0 to 1.5, and then multiplying it by the predicate, but we cannot support it as a basis. Therefore, the above calculation method is not dealt with.

The flow of calculation is (2) → (1) → (3), but if the conditions are not met, it is basically only (1).

Figure 1 shows the flow and conditions of the calculation method.

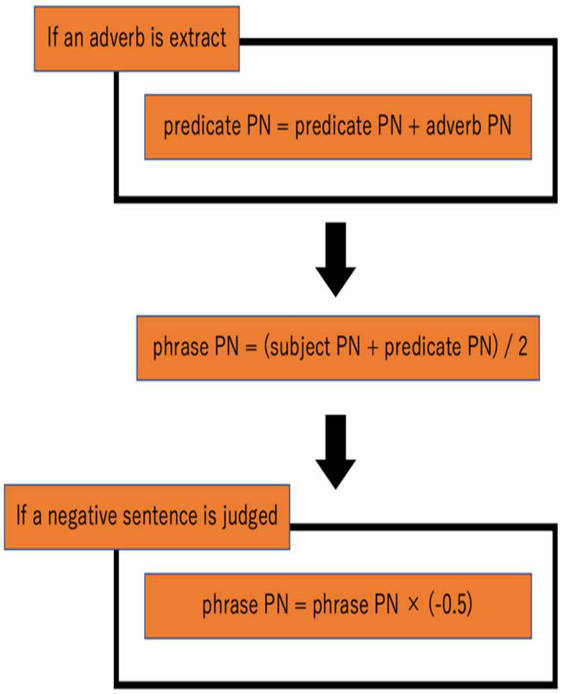

Fig. 1. Overview of How to Calculate Sentiment Analysis

IV. Experiment and Result

We apply the scheme of PN value calculation proposed in Section III to the collected 3,000 tweets based on the location information of a specific location. The subject and predicate are extracted from each tweet parsed structure.

The 674 out of 3,000 tweets contain subject and predicate. The PN value of 142 tweets are positive (PN value is greater than 0) and 467 tweets are negative (PN value is less than 0).

Followings are the four possible patterns of the result of the sentiment analysis.

- The average of subject and predicate is positive.

Example: ABC just before closing is cozy.

PN = 0.4999975((ABC:0[no data] + cozy:0.999995) / 2)

TABLE I. Pattern 1 Words' PN Table

PN	Sentence			
	Subject	***Predicate***	***Adverb***	***not***
value	0	0.999995	0	0

Example: I'm happy with this kind of news in my hometown.

PN = 0.36332175((news:-0.2722275 + happy:0.998871) / 2)

TABLE II. Pattern 2 Words' PN Table

PN	Sentence			
	Subject	***Predicate***	***Adverb***	***not***
value	-0.2722275	0.998871	0	0

- The average of subject and predicate is negative

Example: The end is relatively far.

PN = -0.7641025 ((end:-0.658284 + far:-0.869921) / 2)

TABLE III. Pattern 3 Words' PN Table

PN	Sentence			
	Subject	***Predicate***	***Adverb***	***not***
value	-0.658284	−0.869921	0	0

Example: My legs were unusual and tough yesterday.

PN = -0.793017 ((legs:-0.591837 + tough:-0.994197) / 2)

TABLE IV. Pattern 4 Words' PN Table

PN	Sentence			
	Subject	***Predicate***	***Adverb***	***not***
value	-0.591837	−0.994197	0	0

- Combination of subject, predicate and adverb

Example: There are no chairs in the hall anymore.

PN = -1.16538475 ((chair:-0.654071 + (anymore:-0.676759 + no:-0.9999395)) / 2)

When an adverb is extracted, the total of the predicate and the adverb (2) is treated as one predicate, and the subject and the predicate are calculated (1).

TABLE V. Pattern 5 Words' PN Table

PN	Sentence			
	Subject	***Predicate***	***Adverb***	***not***
value	-0.654071	−0.676759	−0.9999395	0

Example: It's like raindrops falling on the surface of the water.

PN = -1.035225 ((raindrops:-0.630269 + (falling:-0.97775 + like:-0.462431)) / 2)

TABLE VI. Pattern 6 Words' PN Table

PN	Sentence			
	Subject	***Predicate***	***Adverb***	***not***
value	-0.630269	−0.97775	−0.462431	0

- Negative sentence

Example: The south of Koshu Kaido is not called Nakano-dori.

PN = 0.11908558333333334 (((south:-0.47634233333333337 + called:0.0[no data]) / 2) x not: (-0.5))

After calculating the subject and predicate (1), there is a negative sentence judgment, so the phrase is multiplied by *-0.5* (3).

TABLE VII. PATTERN 7 WORDS' PN TABLE

PN	Sentence			
	Subject	***Predicate***	***Adverb***	***not***
value	-0.47634233333333337	0	0	x(−0.5)

Example: Poketalk does not provide translations.

PN = 0.1271405 (((*Poketalk:0.0[no data]* + *provide:-0.508562*) / *2*) x not: (-0.5))

TABLE VIII. PATTERN 8 WORDS' PN TABLE

PN	Sentence			
	Subject	***Predicate***	***Adverb***	***not***
value	0	−0.508562	0	x(−0.5)

Various combinations are possible in the above four patterns.

Figure 2 shows the distribution of numerical values of the sentiment analysis result.

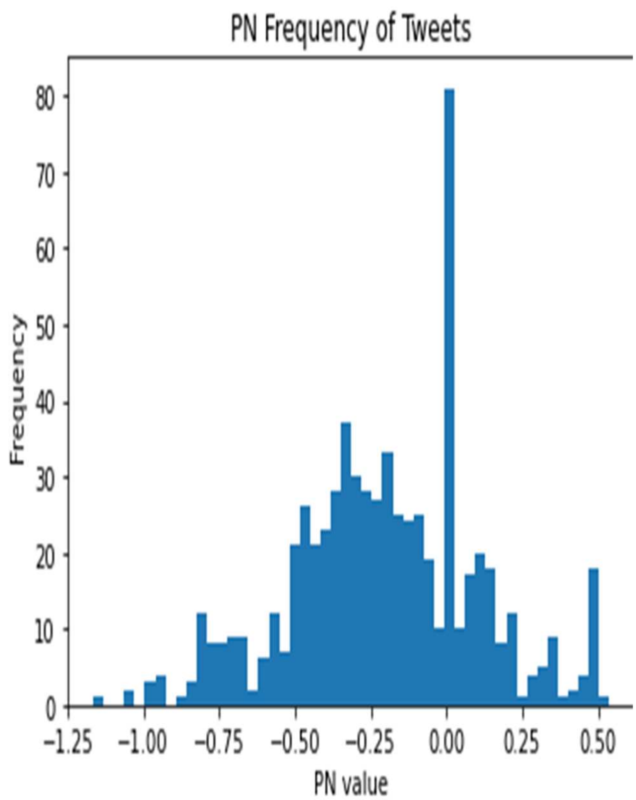

Fig. 2. Distribution of PN values of the collencted 3,000 tweets

From this result, it can be seen that there is basically no clear subject and predicate of the tweet. Also, there is a bias in the distribution of PN values as expected. The negative PN value words occupy the majority in the Word Emotion Polarity Correspondence Table. It is also noticeable that there is a large number of tweets having 0.0 PN values. In the experiment of the limited amount of tweets, we can observe that the PN values and the proposed method of PN calculation can be used to express the location based social interest. Phrase based analysis has been introduced to handle the negation and compound sentences. However, the issue of the coverage of words in the Word Emotion Polarity Correspondence Table is still remained. Since it is linked to location information, from this result, it is considered that the impression of the vicinity is generally negative. Therefore, it can be analyzed that the demand for the place is small, or it may work negatively for itself.

V. CONCLUSION

The study shows the results of sentiment analysis using phrases to solve the problem of misinterpretation in the case of negation and multiple phrases sentence. We propose an efficient scheme of PN value calculation of subject, predicate, adverb, and negator. However, there is still a room for taking into account of other parts of speech and phrase elements the enhance the sentiment analysis of tweets. Since the majority of tweets have unclear subjects and predicates, it can be seen that it is necessary to supplement the subjects and predicates depending on the situation. By finding the PN value for each topic with Latent Dirichlet Allocation (LDA), it is possible to understand what kind of feelings there are for each topic and provide further multifaceted information. In addition, the coverage of sentiment words in the Word Emotion Polarity Correspondence Table can be improved to represent the total sentiment of the location based tweets.

ACKNOWLEDGMENT

This work was partially supported by the Thailand Science Research and Innovation Fundamental Fund, Contract Number TUFF19/2564 and TUFF24/2565.

REFERENCES

[1] TextBlob Sentiment: Calculating Polarity and Subjectivity, June 7, 2015. (https://planspace.org/20150607-textblob_sentiment/)

[2] H. Takamura, T. Inui, M. Okumura, "Extracting Semantic Orientations of Words using Spin Model," Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL2005), pp. 133-140, 2005.

[3] M. Higashiyama, K. Inui, Y. Matsumoto, "Acquisition of Noun Evaluation Polarity Focusing on Predicate Selection Preference," The 14th Annual Conference of The Association for Natural Language Processing, pp.584-587, 18-20 March 2008.

[4] N. Kobayashi, K. Inui, Y. Matsumoto, K. Tateishi and T. Fukushima, "Collecting Evaluative Expressions for Opinion Extraction," The 12th Annual Conference of The Association for Natural Language Processing, pp.203-222 12 July 2005.

[5] A. Esuli, F. Sebastiani, V. G. Moruzzi, "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining," Proceedings of the 5th International Conference on Language Resources and Evaluation, pp. 417–422, 2006.

[6] S. Virach, T. S. Bang, "Sentiment Classification for Hotel Booking Review Based on Sentence Dependency Structure and Sub-Opinion Analysis," IEICE TRANS. INF. & SYST., VOL.E101–D, No.4 April 2018.

[7] H. Sugawara, A. Neviarouskaya, M. Ishizuka, "Affect Extraction from Text in Japanese," The 23rd Annual Conference of the Japanese Society for Artificial Intelligence, 2009.