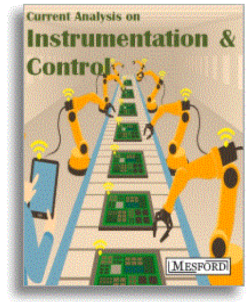

Emotion Recognition for Japanese Short Sentences Including Slangs Based on Bag of Concepts Feature Trained by Large Web Text

Kazuyuki Matsumoto*, Minoru Yoshida and Kenji Kita

Tokushima University, Tokushima, Japan.

Abstract:

The growth of Internet communication sites such as weblogs and social networking sites brought younger people especially in teens and in their 20s to create new words and to use them very often. We prepared an emotion corpus by collecting weblog article texts including new words, analyzed the corpus statistically, and proposed a method to estimate emotions of the texts. Most slang words such as Youth Slang are too ambiguous in sense classification to be registered into the existing dictionaries such as thesaurus. To cope with these words, we created a large scale of Twitter corpus and calculated sense similarities between words. We proposed to convert unknown word to semantic class id so that we might be able to process the words that were not included in the learning data. For calculation similarities between words and converting the word into word cluster id, we used the word embedding algorithms such as word2vec, or GloVe. We defined this method as a method using Bag of Concepts as feature. As a result of the evaluation experiment using several classifiers, the proposed method was proved its robustness for unknown expressions.

Publication History: Received: 28 September 2018 | Revised: 8 November 2018 | Accepted: 14 November 2018

Keywords:
Youth Slang, Unknown Words, Bag of Concepts, Word Embedding, k-nearest neighbor algorithm, Maximum Entropy Method, Unsupervised Clustering.

1. INTRODUCTION

With recent progress of information and communication technology, we can communicate remotely with electrical tools such as e-mail, chat, SNS, BBS and weblog. Because we cannot confirm facial expression or voice tone directly in these text-based communications, we sometimes fail to understand latent implications of the words, especially emotions or feelings. Emoticons or pictographs are effective tools to make text-based communication smoothly. If we use these tools, it becomes easy to convey emotions to other people.

For example, even though emotion is not clearly expressed in the text, a smiley emoticon in the end of the sentence can show that he/she is OK or positive.

Ex.)

1) Next time, I will go out to drink with freshman (^^).

2) Today, I don't go with you (--;).

Of course, such emoticons and pictographs are supplemental. Many complex and varieties of emotions are expressed sensuously in text and they are sometimes hard to be recognized clearly. However, such emoticons and pictographs are just supplementary and there are many complex and rich emotions expressed in text without using them. This emotion recognition becomes even harder with automatic judgment by using evaluation analysis or opinion analysis method in text mining [1],[2].

For example, "Uzakawaii" is a coined word consisting of "Uzai" meaning negative emotions such as annoying or look like fool and "Kawaii" meaning positive emotion such as cute [3]. The word is obviously expressing complex emotion. This word is generally called as "Youth Slang" that is usually used among young Japanese people from teenagers to those in twenties [4].

It is well known that many of the young Japanese people tend to use youth slang so that they can communicate with those who belong to the same community group more smoothly. Recently Internet slang expressions on web are increasingly used among young people who daily use Internet for weblog or BBS.

*Address correspondence to this author at Tokushima University, Tokushima, Japan.; E-mail: matumoto@is.tokushima-u.ac.jp

Mesford Publisher Inc
Office Address: Suite 2205, 350 Webb Drive, Mississauga, ON L5B3W4, Canada; T: +1 (647) 7109849 | E: caic@mesford.ca, contact@mesford.ca, https://mesford.ca/journals/caic/

Many of these expressions are abbreviations of the existing words or neologism and they are difficult to be understood. We thought that it was impossible to ignore Youth Slang for emotion estimation from weblogs because there were many Youth Slang expressing emotion on weblogs. There is no clear boundary between the words used by young people in daily life and words frequently used on Internet, and their definitions are also ambiguous [5].

Therefore, in our study we focused on the newly generated words in the last few years and chose the words thought to be Youth Slang or Internet slang by our own judgment. The chosen words were set as query and the sentences including the query were automatically collected from weblogs. We annotated emotion of the sentences including these expressions and annotated basic forms of these expressions to create a corpus. We statistically analyzed the corpus and proposed an emotion estimation method for the sentences including Youth Slang or Internet slang.

As the follows, the remain of the paper consist of the following sections. Section 2 described about Japanese youth slang. Section 3 describes the youth slang emotion corpus which is built by us. Section 4 introduces the related works, and Section 5 shows our proposed method to estimation emotion based on word concept. Section 6 explains the evaluation experiment, and Section 7 discuss about the experimental results. Finally, Section 8 concludes this paper.

2. YOUTH SLANG

This section describes what Youth Slang is. Youth Slang is defined as slang or jargon used by people from junior high school age to around their thirties. It is typically used to promote communication, amusement or solidarity; to convey ambiguous images; or to hide/alleviate/clarify something.

It is also considered to include specific words or phrases conveying freedom from traditional rules or a sense of amusement [5]. Table **1** indicates typical Youth Slang words appearing from the 1990s to the present.

Table 1. Typical Example of Youth Slang Word.

Itameshi, Shibukaji, Kiyobuta, Uru-uru, KY, Uzai, Kimoi, Kebai, Serebu, Pakuru, Suppin, Owatteru, Takabi-, Shabai, Osare, Kimomen, Ogyaru

As most of these words come and go with the times, existing morphological analysis systems cannot deal with them. As a solution to this problem, the morphological analysis dictionary known as mecab-ipadic-neologd [6] was constructed and released. This dictionary registers new words (neologisms) and is updated regularly. However, if we use this dictionary, processing time will be increased.

It is considered that as words become unnecessary they should be forgotten. The proposed method realizes the versatile emotion estimation method by considering the semantic similarity in the corpus.

In our research, we defined Youth Slang as including Internet slang. To avoid subjective definitions, if the words are included in the "Japanese Slang Dictionary [1]" or "Afureru Shingo [3]", we considered them to be Youth Slang. Also, for the expressions that were found in the general Japanese dictionaries or morphological analysis dictionaries but were not used for their literal meanings and used as slang, we considered such expressions to be Youth Slang. For example, the word "sweets" originally meant sweet things; however, the word also has been used recently as Youth Slang meaning "spoiled."

We manually confirmed what meaning the word was used for. When the word could be judged as being used as Youth Slang, we treated the word as Youth Slang.

3. YOUTH SLANG EMOTION CORPUS

3.1. Annotation

Matsumoto et al. [7] constructed a corpus that consisted of sentences including Japanese Youth Slang. As this corpus has annotated emotion tags in each sentence, it is called the Youth Slang Emotion Corpus (YSEC).

The eight emotions annotated in the corpus are shown in Table **2**. The third column indicates the emotion categories that act as standards for annotating emotion tags. The eight kinds of emotion tags are classified into the emotional polarities of Positive/Negative/Neutral.

Table 2. Basic Emotion Category

Polarity	Emotion Tag	Emotion Categories
Positive	Joy	joy
	Hope	hope
	Love	love, respect
Negative	Hate	hate, anger
	Sorrow	sorrow, shame, regret
	Anxiety	anxiety, fear
Neutral	Surprise	surprise
	Neutral	neutral, reception

Table 3. Basic statistics of YSEC

# of Words	338,779
# of Words per Sentence	16.5
# of Emotion Words	20,672
# of Emotion Words per Sentence	1.01
# of Youth Slang Words	20,147
# of Youth Slang Words per Sentence	0.98
# of Characters	1,805,459
# of Characters per Sentence	88.1
# of Unknown Words	29,830

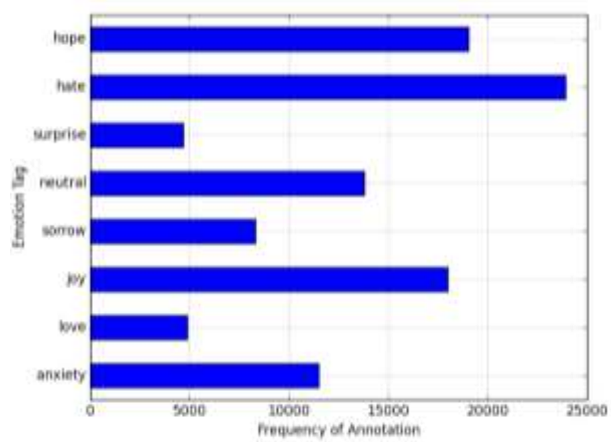

Fig. (1). The distribution of emotion tag.

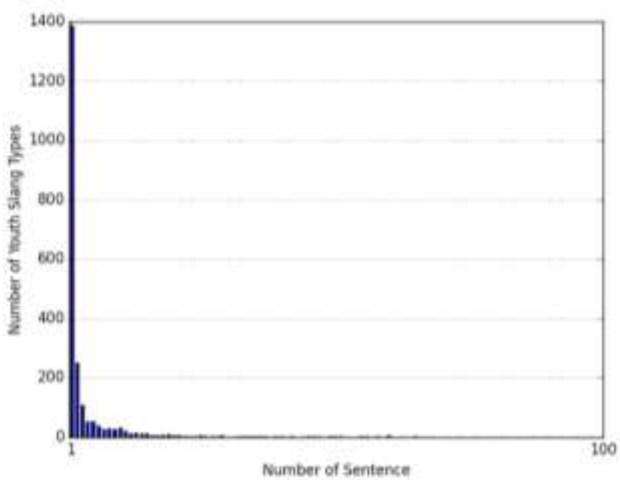

Fig. (2). Distribution of the numbers of example sentences.

3.2. Status of Corpus

This section describes and discusses the results of the statistical corpus analysis. Table **3** shows the basic statistical information of the corpus. Fig. (**1**) shows the distribution of emotion tags annotated to the corpus.

The distribution of the example sentences for each Youth Slang word is indicated as a histogram shown in Fig. (**2**). The figure only includes Youth Slang words with over fifty example sentences.

4. RELATED WORKS

4.1. Emotion Estimation Method

Compared to more formal sentences, such as those in newspapers, the sentences including Youth Slang tend to be ungrammatical in word order, omit particles, or end words in informal ways. In these sentences, the accuracy of preprocessing such as morphological analysis or syntax analysis is expected to be low.

Emura et al. [8] used Support Vector Machine (SVM) with string kernel as a feature for estimating an author's emotion in the Mobile weblogs. By using string kernel, the effects of morphological split error were limited. However, in text classification research, morpheme has been often used as a feature because it is a meaningful minimum unit. Task morphemes should be used as a feature of emotion estimation in particular because word sense is very important.

Ogawa et al. [9] used morpheme unigram for estimating emotion in short Web texts. Mishina et al. [10] obtained approximately 70% accuracy in emotion estimation using the four emotion categories with the method based on the similarity of morpheme n-gram. Matsumoto et al. [11] used the YSEC and evaluated emotion estimation based on Naive Bayes classifier and accumulation method.

However, because these research efforts did not analyze the effect of unknown words such as Youth Slang on emotion estimation, we could not judge whether they were effective for the sentences including Youth Slang.

4.2. Emotion Estimation Using Emotion Word as Feature

Minato et al. [12] studied statistical emotion estimation using emotion word as a feature. They decided the weight of emotion word for each emotion category by using TF-IDF method. However, this method did not consider the co-occurrence with other expressions. Although a certain accuracy was obtained in the case of plain sentences, accuracy was very low in the case of complex sentences.

Matsumoto et al. [13] proposed the emotion recognition method based on Bag of Concepts. Their method used clustering algorithm and sequential of word cluster similarity.

Ren et al. [14] investigated emotion analysis on social big data such as Twitter. Their proposed method used manual annotation corpus and automatic annotation corpus for training of emotion classifier.

4.3. Emotion Estimation Using Sentence Structure Pattern

Emotion estimation based on emotion words has limitations in estimating a sentence's emotion correctly. The reason is that, in most languages, multiple words are combined to express new meanings. If only negative words are added to a sentence, they can express opposite emotions .

Mera et al. [15] proposed an emotion calculation formula. Their method calculated emotion by rating a word's favorability. However, it could not deal with sentences that did not match the case frame. Furthermore, the researchers did not study how to obtain a word's favorability rate for each writer.

The emotion estimation method using sentence pattern for emotion occurrence trigger was proposed by Tokuhisa et al. [16]. As they focused on clearly specified expressions as emotion occurrence triggers, the method would be unsuitable for colloquial expressions with many abbreviations.

4.4. Emotion Estimation based on Deep Neural Networks

Colneriĉ et al. [17] investigated the effectiveness of the word and character-based recurrent convolutional neural networks to estimate emotion from Twitter text. In their study, the deep neural networks based approach could higher performance than that of the Bag of Words based approach.

Abdul-Mageed et al. [18] proposed EmoNet which is an approach of fine-grained emotion detection with Gated Recurrent Neural Networks. Their approach can classify large tweet text into 8 emotion categories by using Gated Recurrent Neural Networks(GRNNs) [19], [20]. They also used distant supervision based on hash tag of tweets. Their method could acquire a superior accuracy of 95.68%.

Matsumoto et al. [21] proposed the emotion estimation from emoticon based on character-level embedding feature and deep convolutional neural networks. Their method could obtain higher accuracy than the baseline methods using n-gram feature and the other machine learning method such as support vector machine, logistic regression, etc. Their method did not target sentence, however, web short text such as tweet often include many non-verbal expression such as emoticon or emoji.

4.5. Existing Research on Youth Slang

Youth Slang often consists of a compound word constructed from existing words. Therefore, Youth Slang fails to be recognized as a word or to be split morphologically. This makes it difficult to distinguish Youth Slang from other words with different meanings and affects negatively emotion estimation. Because most text-based emotion estimation methods use words as features, words fail to be split in the preprocess phase. This will be a problem for emotion estimation.

Harada et al. [22], [23] and Kubomura [24] tried to convert Youth Slang into existing words. In their research, they focused on the deformation types of Youth Slang and their characteristics. They aimed to improve the precision of analysis of sentences including Youth Slang, such as weblog sentences, by converting the slang words to existing words according to a rule.

However, if conversion of Youth Slang into existing words fails, both meaning and emotion of the sentence become very different. Besides, the subtle emotional nuance might be changed by the conversion, even if the meanings are similar.

Matsumoto et al. [25] studied emotion estimation of Youth Slang; however, their method did not treat unknown expressions. Matsumoto et al. [26] also proposed a method to estimate impression of the Youth Slang based on the sensibility of words appearing near the Youth Slang. Ren and Matsumoto et al. [27] used various features and constructed a classifier by machine learning method. They annotated emotion tags to the sentences by using the classifier then semi-automatically constructed a Youth Slang emotion corpus. What is common in these studies is that they chose their methods depending on the elaborately constructed language resources or rules.

5. EMOTION ESTIMATION BASED ON WORD CONCEPT

5.1. Emotion Estimation Method

As in the previous section, we thought that we should avoid emotional change as the result of converting Youth Slang into other known words. On the other hand, emotions are not only expressed by Youth Slang but also by other words and they can be used as features.

Even if any word of the sentence falls out, the sentence's emotion would be expected to change. It would be necessary to use as many words as possible as clues for emotion estimation.

We considered that we could make it easy to process the sentences including unknown expressions by treating the words meaning a certain concept the same as words sharing similar concepts. Therefore, we proposed a method to estimate emotion based on word concept. In the method, words are converted into concepts by referring to context similarity. We thought that this would make emotion estimation possible even for sentences including unknown words such as those in Youth Slang.

5.2. Word Cluster Vector

One of the problems with the existing method using Bag of Words (BoW) as a feature is that it cannot treat unknown words. In many research projects , the problem of insufficient vocabulary was compensated by using a large-scale corpus. New Youth Slang words are being produced almost every day. To treat Youth Slang, it is necessary to update the corpus accordingly; however, this would be impossible if we consider the cost required for tag annotation.

The existing research studied semantic relations between co-occurring words [28], [29], [30]. Kazama et al. [31] tried to obtain the meanings of the words by extracting the Context Similar Words. They created a vector by using the words appearing around the target word as a feature, calculated JS-divergence and extracted the Context Similar Words. However, Context Similar Words indicate the words related to the target word and do not always have the same meaning.

Kuroda et al. [32] analyzed the relations between pairs of words using the Context Similar Words created with the clustering method of Kazama et al. Therefore, the same cluster included the pairs of hypernyms and hyponyms and the pairs of antonyms. It was found that many of the word pairs gathered in the same cluster were related to each other; however, they did not always have the same or similar meaning.

Mikolov [33] judged the positive/negative tone of movie reviews based on a model to predict context words by using the vocabulary around the target word. In this study, they converted the target word into a 400-dimentional vector and classified the reviews by using the vector as a feature.

Then, Joulin et al. [34] proposed the fast learning algorithm of word distributed representation. "fastText" as an implementation of the algorithm is available online [35]. fastText can train on similar strings that have similar vectors to each other by considering character n-grams, this method can robustly differentiate a word notation or unknown expressions.

Pennington et al. [36] proposed a new global log-bilinear regression model that combined two major models of the global matrix factorization and the local context window. This model created a word vector representation with the same or higher accuracy than the Mikolov method.

In the task of emotion classification, it would be inconvenient if the semantically similar words were antonyms. For example, because "dislike" and "like" resemble each other in their usages, they are used as related words although their meanings are opposite to each other. Therefore, it is very important to judge whether the similar words have the same emotion polarity as the target word.

If we check the degree of emotion polarity matching between the similar words and the target word, estimation error might be reduced. Understanding the emotion polarity of unknown words in advance is rare.

In this paper, we made it a priority to increase sentence robustness despite the presence of unknown words. If the words are expressed with vector representations, the sentences can be handled as a set of concepts. Here, the similar concept vectors were aggregated by adding ID to the concepts so that they could be distinguished. The concept set can be expressed simply as the weighted vector summation or as the weighted concept vector. As indicated in Fig. **3**, our proposed method can abstract a sentence better than the existing methods using words or word n-grams as concepts. Therefore, it would be easy to process the sentences including unknown expressions.

To analyze how much the words that have similar meanings but different emotion polarities influence emotion estimation, we clustered the words based on the word context vector and calculated the emotion polarity match rate in the same cluster.

We prepared a Tweet corpus by collecting Tweets based on Youth Slang and emotion expression lists. By setting window as 10 and size as 200, word2vec [37], [38] was trained to create word context vectors. As a clustering tool, bayon [39] was used. The number of clusters was set to 30,000 according to the repeated bisection method [40]. We calculated the sum of the emotion polarity values of the words in each cluster, then calculated how many of them had the same emotion polarity as the target word.

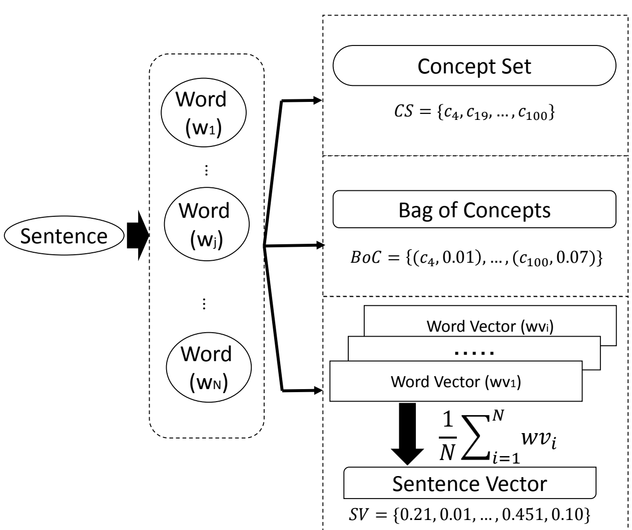

Fig. (3). Concept representation of a sentence.

Equation 1 indicates calculation of emotion polarity match rate in a cluster. Equation 2 calculates the polarity of the cluster, which is the summation of the emotion polarities for words $w_{ij}*C_i$ that are in the cluster C_i and have emotion polarities.

$Match(p(C_i), sgn(pn_{ij}))$ is a function that returns 1 if the polarity of cluster C_i matches with the polarity of the word w_{ij} in the cluster and returns 0 if there is no match. $|C_i|$ indicates the number of the words having emotion polarity in cluster C_i.

$$emr_i = \frac{1}{|C_i|} \times \sum \; Match\left(p(C_i), sgn(pn_{ij})\right) \quad (1)$$

$$p(C_i) = sgn\left\{\sum_{w_{ij} \in C_i} pn_{ij}\right\} \quad (2)$$

Emotion polarity is annotated to the words by referring to the Japanese appraisal evaluation expression dictionary [41]. The resulting emotion polarity match rates calculated in each corpus are shown in Table **4**.

The corpus type 'em' indicates the Tweet corpus collected based on emotion expressions, and 'ys' indicates the Tweet corpus collected based on Youth Slang. The table shows that approximately 90% of the emotion polarity match rates were obtained in many clusters.

Table IV. Emotion polarity match rate for each number of cluster and each corpus.

Corpus	Number of Clusters	emr	(Total Match Rate / Cluster no.)
em	5000	0.81	(2243.8 / 2773)
em	10000	0.86	(3419.9 / 3958)
em	15000	0.89	(4128.4 / 4633)
em	20000	0.91	(4615.3 / 5090)
em	25000	0.92	(5012.6 / 5446)
em	30000	0.93	(5292.3 / 5694)
ys	5000	0.84	(1779.5 / 2109)
ys	10000	0.88	(2411.8 / 2740)
ys	15000	0.90	(2781.3 / 3085)
ys	20000	0.92	(3054.5 / 3332)
ys	25000	0.93	(3273.6 / 3523)
ys	30000	0.94	(3415.4 / 3647)

That means, when there is a word having emotion polarity in a cluster, the words having different emotion polarity would be rarely included in the cluster.

We conducted unsupervised clustering on the context vectors so that the words could be converted into cluster IDs. By converting a sentence into the fixed dimension binary vector (vector's dimension is the number of clusters), it would become possible to simplify the similarity calculation between the sentences and to strengthen the response to the words that are not included in the training data. Fig. (**4**) shows the training flow of the word cluster vector from the Tweet corpus.

Fig. (**5**) shows the flow of emotion estimation from the inputted sentences by using the word cluster vector. The system converts the list of words extracted from the inputted sentence into the word cluster id vector by referring to the word cluster database.

The obtained vector is compared with the word cluster ID vector obtained from the YSEC. Finally, the system outputs the emotion tag of the similar sentence. We proposed the following four classification methods in this paper:

- Classification by k-nearest neighbor method based on similarity between Concept Sets
- Classification by Maximum Entropy Method (MEM) by using Concept Set as feature
- Classification by MEM by using Bag of Concepts (BoC) as feature
- Classification by MEM by using Sentence Vector as feature

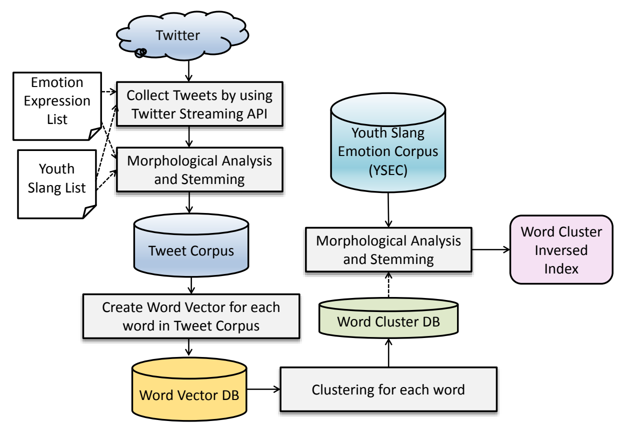

Fig. (4). Flow of training of word cluster vector.

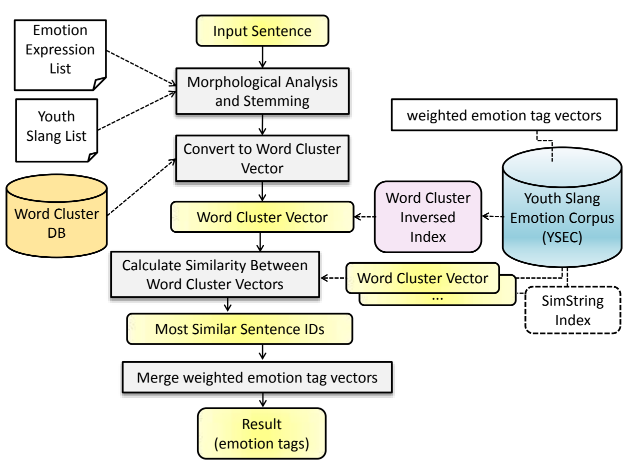

Fig. (5). Estimation flow using k-NN method.

The definition of the BoC is as follows. In this paper, the BoC is a feature vector created by calculating the sum of the affiliation degrees for the word concepts appearing in a sentence.

The word concepts are clustered by using a context vector obtained from the corpus. By doing this, we aimed to improve the decrease in accuracy caused by the word having multiple registered concepts in the existing concept dictionaries.

The calculation cost becomes high when similarity is calculated for all data. Therefore, the calculation candidates were obtained by using high-speed search for similar letter strings. For this high-speed search for similar strings, we used an open source library called SimString that was developed by Okazaki et al. [42].

6. EXPERIMENT

6.1. Experimental Data

Table **5** shows the details of the test data and the training data that we used in the experiment. Because we conducted the experiment with 10-hold cross validation, the training data included the duplicate sentences and they were counted.

6.2. Experimental Conditions

In the evaluation experiment, we compared the following five methods:

1. Baseline method (maximum entropy classifier based on morpheme 1-gram: BoW)
2. k-NN classifier based on Concept Set feature
3. Maximum entropy classifier based on Concept Set feature
4. Maximum entropy classifier based on BoC feature
5. Maximum entropy classifier based on Sentence Vector feature

Methods 3 and 4 can find optimal training parameters for each condition because the methods depend on the algorithm of word vector generation or the number of clusters. In the k-NN method, similarity degree was calculated by using Concept Set as a feature because we valued high-speed performance.

K-nearest neighbor algorithm takes a kind of lazy learning based on instance. Although the method is simple, it promises a certain level of accuracy for the high quality and exhaustive database, and therefore is used for various classification tasks.

Because the proposed method conducts clustering on the word context vector, it can reduce the dimension of the feature vector to learn to some degree. However, over-reducing the dimension can result in decrease of accuracy.

In this research, after obtaining the cluster vector lists having similarity degrees higher than a certain threshold level as candidates, we conducted detailed similarity calculations. The cluster string is a list that lists the cluster IDs in order of appearance in the sentence. We used SimString to extract candidates.

The cluster strings were listed in order of appearance to avoid the local similar string and global similar string from being handled as candidates of the same level. Because the inputted sentences are unknown data, they will never match with the training data. Therefore, we used cosine similarity and set similarity threshold as 0.2 or more for candidate extraction. To decide final rank, we used cosine similarity among the concept set vectors.

As the evaluation metrics, we measured the Recall, Precision, and F1-Score of the proposed method and the comparison methods by using a 10-fold cross-validation approach to evaluate the test data and the training data. The calculation formulas are shown in Equation 3, 4, and 5.

Table 5. Experimental Data.

Emotion	# of Test Data	# of Training Data
Anxiety	392	5,371
Love	394	5,302
Sorrow	420	5,627
Surprise	459	6,032
Joy	525	6,969
Hate	487	6,331
Hope	449	6,019

Table 6. Result of the Baseline Method.

Emotion	Recall	Precision	F1-score
Anxiety	0.183	0.232	0.204
Love	0.357	0.398	0.376
Sorrow	0.232	0.273	0.251
Surprise	0.333	0.371	0.351
Joy	0.372	0.335	0.353
Hate	0.320	0.269	0.292
Hope	0.247	0.195	0.218

$$\mathrm{Recall}_e = \frac{tp_e}{tp_e + fn_e} \quad (3)$$

$$\mathrm{Precision}_e = \frac{tp_e}{tp_e + fp_e} \quad (4)$$

$$\text{F1-score} = 2 \cdot \frac{\mathrm{Recall}_e \cdot \mathrm{Precision}_e}{\mathrm{Recall}_e + \mathrm{Precision}_e} \quad (5)$$

In Equation 3 and 4, tp_e indicates true positive which is an outcome where the system correctly estimates category e. In Equation 3, fn_e indicates false negative which is an outcome where the system incorrectly estimates category e. In Equation 4, fp_e indicates false positive which is an outcome where the system incorrectly estimates the other category as category e. Recall, Precision and F1-score are often used for text category classification task or information retrieval task [43].

In this evaluation experiment, for context vector clustering, we tried the six cluster numbers of 5000, 10000, 15000, 20000, 25000 and 30000 and then varied the k-value of k-NN from 1 to 100 gradually.

In the experiment, we also used MEM as classification method and the L-BFGS [44] as machine learning algorithm. Classias was used as machine learning program. We used word 1-gram, Concept Set, BoC, and Sentence Vector as features for training with MEM. Sentence Vector has the linear summation of word vector learned by Word2Vec or GloVe as a feature. In the BoC, we calculated summation of the affiliation degrees of each word in the sentence to each concept cluster and then created a feature vector by weighting concept ID with that summation of the affiliation degrees.

7. RESULTS AND DISCUSSIONS

7.1. Result of Baseline Method

This section describes the experimental result of the baseline method using morpheme n-gram as a feature. The experimental result of Baseline is shown in Table **6**. Recall, Precision and F1-Score were calculated according to each emotion category. In total, approximately 29.5% accuracy was obtained, and the results showed that the emotion of "love" obtained highest F1-Score of approximately 0.376.

7.2. Result of Proposed Method

Fig. (**6**, **7**, **8**, **9**) show the transition of accuracy in the evaluation experiment of the proposed method using k-NN. Figs. (**10**, **11**, **12**, **13** show accuracies with the MEM using Concept Set and BoC as features.

From this figure, we could not find explicit difference in accuracies between Concept Set and BoC. However, there were some differences depending on the corpus. The accuracies with ys corpus were always higher in any number of clusters. Higher accuracies were stably obtained when Word2Vec was used for vector generation. k-NN method requires higher cost than MEM; however, it produces stable accuracy by using appropriate parameters. This method is effective in certain ways, in that it can respond to the change of data if the SimString index is created quickly.

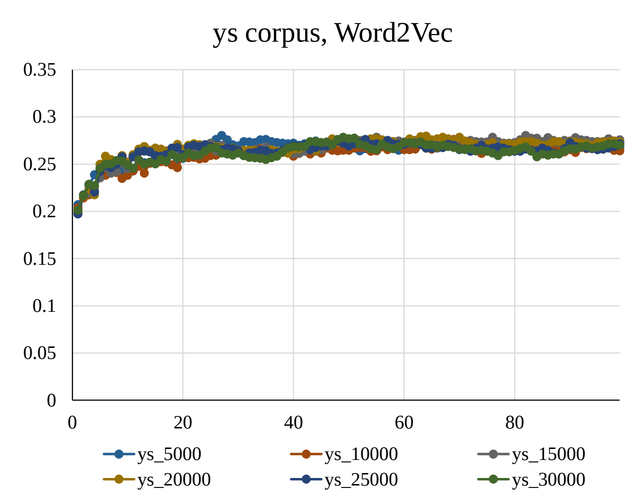

Fig. (6). Result of k-NN using BoC as feature(ys, Word2Vec).

Table **7** and **8** show the Recall, Precision and F1-Score obtained for each emotion when Sentence Vector was used as a feature. The result did not show big difference in F1-Score depending on the corpus even though there was a difference depending on emotions. That means there were no big differences in the generated vectors.

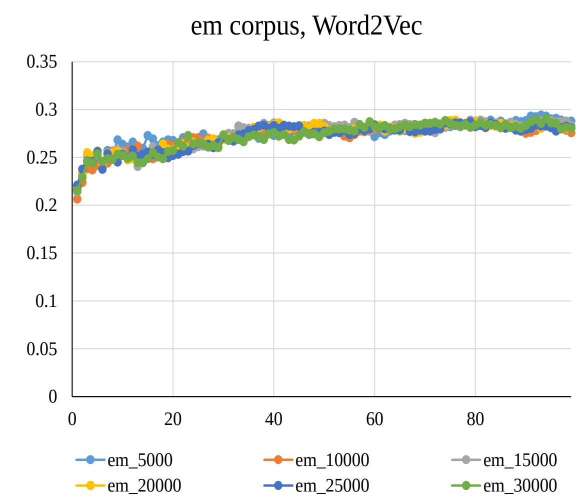

Fig. (7). Result of k-NN using BoC as feature(em, Word2Vec).

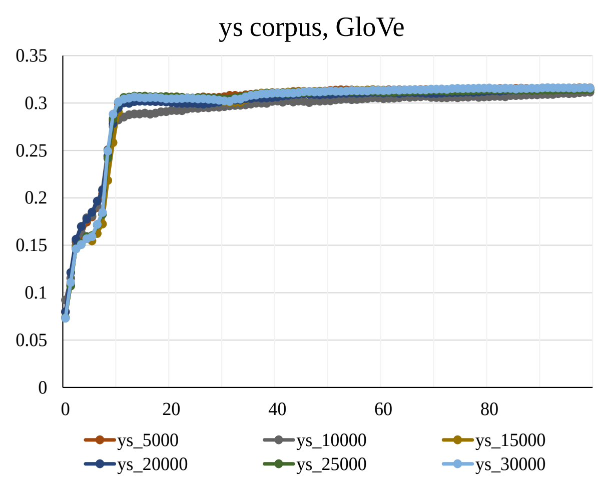

Fig. (8). Result of k-NN using BoC as feature(ys, GloVe).

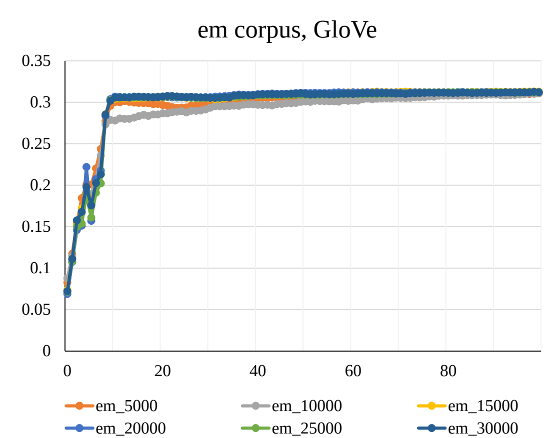

Fig. (9). Result of k-NN using BoC as feature(em, GloVe).

Table 7. Recall(R), Precision(P) and F1-Score(F) for each emotion when the method based on Sentence Vector feature was used (using Word2Vec).

Emotion	ys			em		
	P	R	F	P	R	F
Anxiety	0.40	0.40	0.40	0.36	0.38	0.37
Love	0.58	0.56	0.57	0.54	0.51	0.52
Sorrow	0.43	0.42	0.42	0.38	0.38	0.38
Surprise	0.46	0.45	0.46	0.42	0.42	0.42
Joy	0.44	0.45	0.45	0.42	0.41	0.41
Hate	0.41	0.42	0.41	0.41	0.40	0.40
Hope	0.44	0.43	0.44	0.39	0.38	0.39

Table 8. Recall(R), Precision(P) and F1-Score(F) for each emotion when the method based on Sentence Vector feature was used (using GloVe).

Emotion	ys			em		
	P	R	F	P	R	F
Anxiety	0.21	0.28	0.24	0.29	0.35	0.32
Love	0.50	0.45	0.47	0.52	0.49	0.51
Sorrow	0.38	0.38	0.38	0.41	0.38	0.39
Surprise	0.47	0.45	0.46	0.47	0.44	0.45
Joy	0.40	0.42	0.41	0.42	0.44	0.43
Hate	0.41	0.36	0.38	0.39	0.38	0.38
Hope	0.40	0.42	0.41	0.41	0.44	0.42

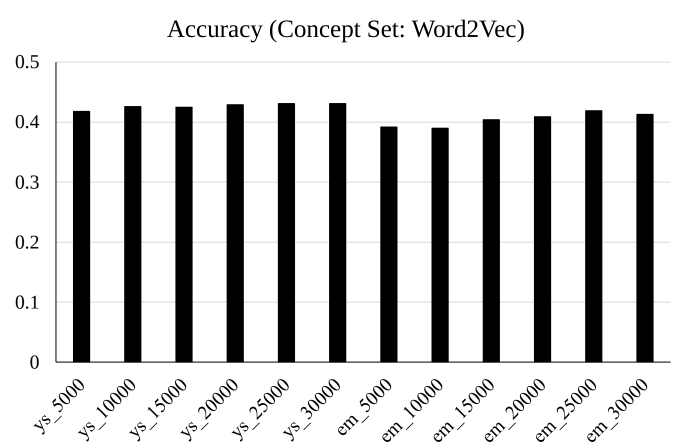

Fig. (10). Result of the maximum entropy based method using Concept Set and Word2Vec.

Table 9. Relation of successful/failed emotion estimation and number of unknown words.

Method		N_s	N_f	$N_s/(N_s + N_f)$
Baseline		159	536	0.229
Concept Set (ys:30000)	Wor2Vec	212	483	0.305
	GloVe	177	518	0.255
BoC (ys:30000)	Wor2Vec	220	475	0.317
	GloVe	194	501	0.279
Sentence Vector (ys)	Wor2Vec	313	382	0.450
	GloVe	217	478	0.312

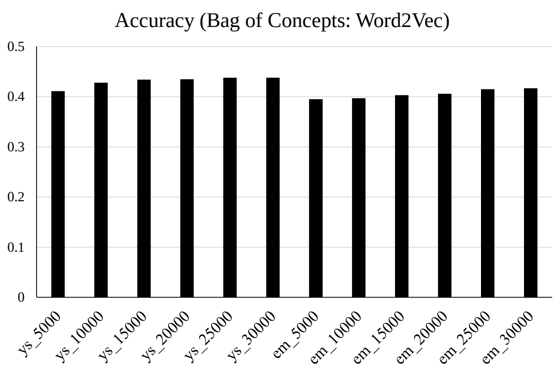

Fig. (11). Result of the maximum entropy based method using Bag of Concepts and Word2Vec.

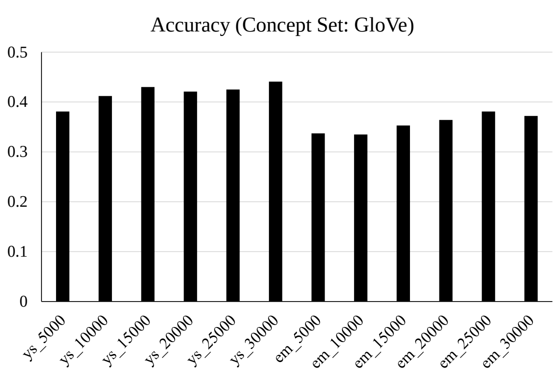

Fig. (12). Result of the maximum entropy based method using Concept Set and GloVe.

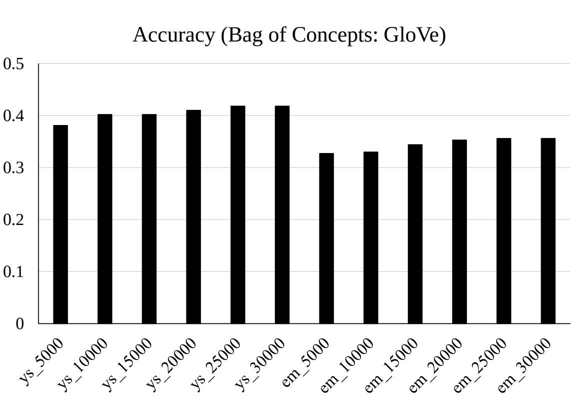

Fig. (13). Result of the maximum entropy based method using Bag of Concepts and GloVe.

7.3. Discussions

We discuss the problems of the proposed method. The larger the vector dimension becomes and the wider the context window size becomes, the more detailed difference between the words would be obtained. However, the most appropriate parameter can be obtained only from empirical approaches.

The corpus size and the queries used for collecting tweets were different in the two tweet corpora that were used in the experiment. We consider that the difference of the quality rather than the size resulted in the difference. The ys corpus had the sentences consisting of almost only with slang words.

For example, a sentence of "Bucchake, majikichi!" consists of the two slang words of "Bucchake" and "Majikichi" and it is quite enough to convey the intent of the sentence with the two slang words.

Such tendency suggests that similar slangs should be easier to be classified into the same cluster by using ys corpus. Therefore, we thought that the high accuracies were obtained in the sentences including slang words. In addition, more trials would be required to acquire the appropriate number of clusters.

CONCLUSIONS

In this paper, we constructed a corpus consisting of the sentences including Youth Slang to estimate emotion from the sentences including Youth Slang.

We proposed a method using Bag of Concepts as feature and evaluated the accuracy by the evaluation experiment comparing the proposed method and the existing method based on Bag of Words as feature.

As the result, it was found that the proposed method could obtain approx. 44% accuracy in emotion estimation by using Bag of Concepts as feature. The proposed k-NN method could not obtain such high accuracy, however, we considered that the accuracy would be improved by using the centroid vector of high-purity cluster. We also could find the effectiveness of the proposed method for unknown expressions.

The method using sentence vector obtained same or higher accuracies compared to the method based on Bag of Concepts, however, the calculation of vector summation might be affected by sentence length. In future, we would like to improve the proposed method by analyzing the relation between the sentence length and the estimation accuracy. Also, for comparing with our proposed method, we would like to use deep neural networks such as character-level recurrent neural networks, convolutional neural networks, and long short-term memory, etc. to generate sentence or slang representation for emotion classification.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number JP15K16077.

REFERENCES

[1]. Turney PD. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02); 2002 July; Philadelphia, USA. ACL 2002.

[2]. Otsuka Y, Inui T, Okumura M. Opinion analysis engine. Coronasha, (in Japanese) , 2007.

[3]. Kitahara Y. Afureru shingo. Taishukan Shoten, (in Japanese), 2009.

[4]. Yamaguchi N. Wakamonokotoba ni mimiwo sumaseba. Kodansha, (in Japanese), 2007.

[5]. Yonekawa A. Study on young word. Meiji Shoin, (in Japanese), (1998).

[6]. Toshinori S. Neologism dictionary based on the language resources on the web for Mecab, https://github.com/neologd/mecab-ipadic-neologd, 2015.

[7]. Matsumoto K, Kita K, Ren F. Emotion estimation from sentence using relation between Japanese slangs and emotion expressions. Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation;2012 Nov 7-10; Bali, Indonesia. ACL 2012.

[8]. Emura K, Yasugi M, Miyazaki S, Kuboyama T. SVM-based emotion extraction from moblog text. IEICE Technical Report 2007; 106(473): 61–6.

[9]. Ogawa T, Matsumoto K, Ren F. About emotion estimation of 'Emonyu' short sentence. IPSJ SIG Technical Report, 2010; 2010-NL-195, (29): 1–6.

[10]. Mishina K, Tsuchiya S, Ren F. Comparison between the human emotion transfer ratio and the similarities of emotion. Proceedings of the 1st International Conference on Advanced Intelligence (ICAI-08); Oct 19-22; Beijing, China.

[11]. Matsumoto K, Konishi Y, Sayama H, Ren F. Wakamono kotoba emotion corpus and its application for emotion estimation. Proceedings of the 2010 International Conference on Advanced Intelligence, 2010 Aug 20-23; Beijing, China.

[12]. Minato J, Matsumoto K, Ren F, Tsuchiya S, Kuroiwa S. Evaluation of emotion estimation methods based on statistic features of emotion tagged corpus. International Journal of Innovative Computing, Information and Control 2008; 4(8): 1931-41.

[13]. Matsumoto K, Yoshida M, Xiao Q, Xin L, Kita K. Emotion recognition for sentences with unknown expressions based on semantic similarity by using bag of concepts. Proceedings of the 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD'15), 2015 Aug 15-17; Zhangjiajie, China. IEEE 2015.

[14]. Ren F, Matsumoto K. Emotion analysis on social big data. ZTE Communications 2017; 15(S2): 30-7.

[15]. Mera K, Ichimura T. Emotion generating method on human-computer interfaces. In: Seppo J. Ovaska, Eds. Computationally intelligent hybrid Systems: the fusion of soft computing and hard computing. Wiley-IEEE Press 2005; pp. 277–315.

[16]. Tokuhisa R, Inui K, Matsumoto Y. Emotion classification using massive examples extracted from the web. IPSJ-Journal, 2009; 50(4): 1365–74.

[17]. Colneriĉ N, Demsar J. Emotion recognition on Twitter: comparative study and training a unison model. IEEE Transactions on Affective Computing, 2018; DOI: 10.1109/TAFFC.2018.2807817.

[18]. Abdul-Mageed M, Ungar L, EmoNet: fine-grained emotion detection with gated recurrent neural networks. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017 July 30-Aug 4; Vancouver, Canada. ACL 2017. DOI: 10.18653/v1/P17-1067.

[19]. Cho K, Merrienboer BV, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014 Oct 25-29; Doha, Qatar. ACL 2014.

[20]. Chung J, Gulcehre C, Cho K, Bengio Y. Gated Feedback recurrent neural networks. Proceeding of the 32nd International Conference on International Conference on Machine Learning, 2015 July 6-11; Lille, France. ACM 2015.

[21]. Matsumoto K, Fujisawa A, Yoshida M, Kita K, Emotion recognition of emoticon based on character embedding, Journal of Software 2017; 12(11): 849-57.

[22]. Harada T, Kameda H. Younger's word processing methods and their evaluation. Technical report of IEICE TL2002-33, (in Japanese), 2002; pp. 5–6.

[23]. Harada T, Yamamoto Y, Kubomura C, Sasaki Y, Kameda H. Evaluation of younger's word processing system. Proceedings of the IEICE General Conference, 2006, Mar 24-27; Tokyo, Japan. (in Japanese), 2006.

[24]. Kubomura C, Harada T, Sasaki Y, Yamamoto Y, Kameda H. An evaluation method of a younger's word processing system with use of blog articles. Technical report of IEICE. Thought and language, 2006; 105(613): 165–169.

[25]. Matsumoto K, Ren F. Construction of wakamono kotoba emotion dictionary and its application. Proceedings of the 12th International Conference on Intelligent Text Processing and Computational Linguistics, 2011 Feb 20-26; Tokyo, Japan. Springer 2011.

[26]. Matsumoto K, Yoshida M, Kita K. Sensibility estimation method for youth slang by using sensibility co-occurrence feature vector obtained from microblog. Proceedings of the 2015 IEEE International Conference on Computer and Communications (ICCC2015), 2015 Oct 10-11; Chengdu, China. IEEE 2015.

[27]. Ren F, Matsumoto K. Semi-automatic creation of youth slang corpus and its application to affective computing. IEEE Transactions on Affective Computing 2016; 7(2): pp. 1-16, DOI: 10.1109/TAFFC.2015.2457915.

[28]. Pantel P, Pennacchiotti M. Espresso: leveraging generic patterns for automatically harvesting semantic relations. Proceedings of the 21st COLING and 44th ACL(COLING-ACL-06), 2006 July 17-18; Sydney, Australia. ACM 2006.

[29]. Saeger SD, Torisawa K, Kazama J, Kuroda K, Murata M. Large scale relation acquisition using class dependent patterns. Proceedings of the 9th IEEE International Conference on Data Mining, 2009 Dec 6-9; Miami, Florida. IEEE 2009.

[30]. Turney PD. A uniform approach to analogies, synonyms, antonyms, and associations. Proceedings of the The 22nd International Conference on Computational Linguistics, Manchester, 2008 Aug 18-22; Manchester, UK. ACL 2008.

[31]. Kazama J, Torisawa K. Inducing gazetteers for named entity recognition by largescale clustering of dependency relations. Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics, 2008 Jun 15-20; Ohio, USA. ACL 2008.

[32]. Kuroda K, Kazama J, Torisawa K, A Look Inside the Distributionally Similar Terms, In: Proceedings of the second workshop on NLP challenges in the information expansion era (NLPIX2010), 2010; pp. 40-49.

[33]. Le Q, Mikolov T. Distributed representations of sentences and documents. Proceedings of the 31st International Conference on Machine Learning, 2014 Jun 21-26; Beijing, China.

[34]. Joulin A, Grave E, Bojanowski P, Mikolov T. Bag of tricks for efficient text classification. Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, 2017 Apr 3-7; Valencia, Spain. ACL 2017.

[35]. fastText: Available from: https://github.com/facebookresearch/fastText

[36]. J. Pennington, R. Socher, and C. D. Manning. GloVe: global vectors for word representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014 Oct 25-29; Doha, Qatar. ACL 2014.

[37]. Word2Vec: https://github.com/dav/word2vec

[38]. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in Vector Space, Available from: https://arxiv.org/abs/1301.3781

[39]. Bayon: https://github.com/fujimizu/bayon

[40]. Zhao Y, Karypis G. Comparison of agglomerative and partitional document clustering algorithms. Technical report, Department of Computer Science, University of Minnesota, Minneapolis, 2002; MN 55455.

[41]. M. Sano. Reconstructing english system of attitude for the application to Japanese: an exploration for the construction of a Japanese dictionary of appraisal. Proceeding of the 38th International Systemic Functional Congress, 2011 Jul 25-29; Lisbon, Portugal.

[42]. Okazaki N, Tsujii J. A simple and fast algorithm for approximate String matching with set similarity. The journal of natural language processing 2011; 18(2): 89–117.

[43]. Manning CD, Raghavan P, Schütze H. Introduction to information retrieval. Cambridge University Press. 2008.

[44]. Nocedal J. Updating quasi-Newton matrices with limited storage. Mathematics of Computation 1980; 35(151) 773–82.