Adjective-Based Estimation of Short Sentence's Impression

Nguyen Thi Thu An[1], Masafumi Hagiwara[2]

[1] *Graduation School of Science and Technology, Keio University, Yokohama, Japan, annguyen@soft.ics.keio.ac.jp*

[2] *Graduation School of Science and Technology, Keio University, Yokohama, Japan, hagiwara@soft.ics.keio.ac.jp*

Abstract: This paper proposes a new method to estimate impression of short sentences considering adjectives. In the proposed system, first, an input sentence is analyzed and preprocessed to obtain keywords. Next, adjectives are taken out from the data which is queried from Google N-gram corpus using keywords-based templates. The semantic similarity scores between the keywords and adjectives are then computed by combining several computational measurements such as Jaccard coefficient, Dice coefficient, Overlap coefficient, and Pointwise mutual information. In the next step, the library sentiment of patterns.en - natural language processing toolkit is utilized to check the sentiment polarity (positive or negative) of adjectives and sentences. Finally, adjectives are ranked and top na adjectives (in this paper na is 5) are chosen according to the estimated values. We carried out subjective experiments and obtained fairly good results. For example, when the input sentence is "It is snowy", selected adjectives and their scores are: white (0.70), light (0.49), cold (0.43), solid (0.38) and scenic (0.37).

Keywords: Impression, polarity, relatedness, semantic similarity.

1. INTRODUCTION

Estimation of semantic similarity between words or some entities is important. It is one of the essential and fundamental technologies for information retrieval and natural language processing. It has a large range of applications, such as word sense disambiguation (Linlin Li, Benjamin Roth, and Caroline, 2010), information retrieval (Saini, & Sharma, & Gupta, 2011), paraphrase recognition (Prodromos, 2009), text summarization and annotation (Ramiz, 2009), and lexical selection (Jian, Lujun, Yang, Hua-Jun, Hua, Qiang & Zheng, 2008). Several measures have been proposed to compute the relatedness scores. These scores can show the similarity between words, a word and its relation. There are semantic relatedness measurements: Ref. (Alexander & Graeme, 2006) introduced the second order co-occurrence pointwise mutual information as a measure of semantic similarity using the British National Corpus (BNC). Ref. (Mehran Sahami & Timothy, 2006) evaluated knowledge-based measures of word relatedness using *WordNet* as their central resource. Some of the other researchers define the semantic relatedness between the words using Web (Huirong, Pengbin, Baocai, Mengduo & Yanyan, 2011), (Danushka, Yutaka, and Mitsuru, 2011), (Michael & Simone, 2006), (Lun, Yong and Hsin, 2007). Most of the researches, however, just focused on the relatedness between words (noun), a word (a noun) and concepts or entities, which seems not sufficient if we consider realistic applications. For examples: *"summer comes"* associates that the temperature is high, whereas *"summer passes"* means the temperature is not high. Therefore, the estimation of the relatedness between sentence and adjective is one of the most important researches for new trend of *Kansei* research.

In many languages, adjectives are one of the most important elements. They are the best indicator of subjectivity. Usage of adjectives is a popular way to express our feeling, and our impression about something or the quality of any facts or events. They make the content of what speakers tell become more visual and vivid. Also, they are used for descriptive talk or writing.

Sentiment analysis and opinion mining refer to applications and researches related to usage of adjectives. Their objective is to identify and to extract the subjective information in source materials. It has received considerable attention in the research community. The task of polarity and emotion identification may contribute to a broad variety of possible applications such as recommender systems, the collection of opinions in product reviews, in financial news and also in the domain of human computer interaction. The semantic orientation can be analyzed from various levels such as words, sentences, phrases or on entire documents (Subrahmanian & Reforgiato, 2008), (Samaneh & Fred, 2010), (Tim, 2011). These methods extract adjectives and their frequencies from the given reviews, and then it is able to predict the polarity of each adjective using the learned classifier, and classify the review based on the polarity of the adjectives. Current researches have obtained many achievements, however, they are just limited in such areas: judgment or evaluation, whereas they are able to apply and extend to the others, impression estimation in particular.

In this paper, we propose a new method that adopts the association measurement and the sentiment analysis as the main tasks. One of the advantages to compute the relatedness strength between words is that system can express the impression of a given sentence using adjective. The main contribution of this method is to propose a new concept of semantic association - adjective and sentence, and to propose an impression estimation system.

2. PROPOSED SYSTEM

Before detailed explanation, we give a brief about the process of the proposed system. Fig.1 illustrates flow of the process. There are 3 main steps: keywords extraction, adjectives collection, and their ranking. First, preprocessing is carried out to extract keywords from the input sentence. In this step, we group the words by each part of speech and select one word as a keyword of each group. Selected keyword candidates are then used to query the *N*-gram utterances (N = 5 in this paper) which are used in the next step - adjective candidate extraction. The dataset for the proposed system is a collection of keywords and the candidate adjectives queries based on Google *N*-gram. Then, we apply association measures to compute the correlation between adjective and keywords. We check the polarity orientation of the adjectives and the input using `Sentiwordnet` (Kerstin, 2008), and `Patterns.en` library. Finally, the top n_a adjectives having the highest similarity scores and the same orientation with the input sentence are displayed.

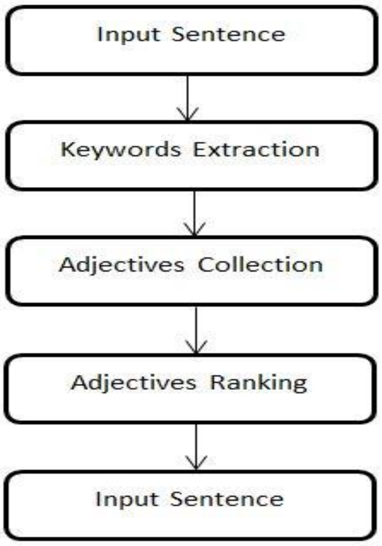

Figure 1: Overview of the proposed system

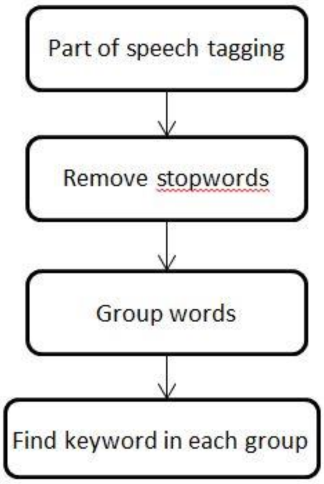

Figure 2: Overview of Keyword Extraction

To compute the similarity between each collected adjective and keywords, there are two cases we need to deal with. In the first case - only a single keyword, the relatedness score between keyword and adjective is computed using corpus-based association measures: Jaccard, Dice, Overlap and PMI. In the second case when there are plural number of keywords, the multivariate association measures (Tim, 2011) and Dice for multivariate are used to compute the relatedness score between the keywords and the adjectives.

The sentiment checking module of `Patterns.en` and `Sentiwordnet` (Kerstin, 2008) are used to check the polarity of adjective and input. The following sections give a step-by-step explanation of the proposed system using the following examples to easy understanding.

2.1. Keywords Extraction

Keywords extraction is detailed in Fig.2. A sentence is preprocessed based on part of speech (POS) tagging and stopwords removal. For the task of stopwords removal, *Wordnet* corpus of stopwords is used to filter words, little lexical content, out of the given sentence. Then, the rest of the words are grouped by POS tagging. Each word in each category queries unigram frequency based on Google *N*-gram corpus. For the case when these words have higher frequencies, it can be considered as common and not important ones. Therefore, we regard the word in each category with minimum frequency as a keyword of the group.

Here, we explain the above keywords extraction steps using an example when the input is *"A lot of trees were blown down in the storms"*.

- Step1: Part of speech tagging:

"A" and "*the"* appear as determines of sentence; "*lot, trees, storms* are nouns; *blown*, *were* are verbs; *of*, *in* are prepositions; *down* is adverb.

Words after removing stopwords: *trees, blown, storms*.

- Step 2: Frequencies of group and words are queried:
  - Nouns:
  - Verbs: *blown*
    - *trees:* 54,000,000.
    - *storms:* 12,530,000.
  - Verbs: *blown*
- Step 3: Keywords decision:
  - Key noun: *storms*
  - Key verb: *blown*

Result: Keywords are *storms, blown*.

In our research, inputs are short sentences, so obtaining keywords belongs to one of the following two cases:

- Case 1: A word playing the role of keyword. That word might be a noun, a verb or an adjective.
- Case 2: A pair of words playing a role of keyword. These two words are noun and verb.

2.2. Adjective Collection

This step aims to have a dataset of *N*-adjectives (N_a) which are related most to the obtained keywords. The keywords are then used to create templates to query Google *N*-gram. The resulting queries are continually processed. Finally, all of the adjectives are collected. Fig. 3 shows the overview of adjective collection.

Later tables show some examples of the templates created from keywords. The order of the words in the query is free. The result is a list of *N*-gram chunks. To have an effective retrieval, keywords are changed to many forms as what we mentioned - form of noun, verb and adjective. The following explanations give the templates for each situation.

In the case keyword is a noun, templates are illustrated in Table 1.

Input: *sunrise*

Keyword: *sunrise*

After using these templates to query, the *N*-gram chunks are preprocessed (tokenizing, Part of speech tagging), and then adjectives are extracted and saved in a list as shown in Table 2.

Similarly, for the case of verb and the case of adjective playing the keyword role, the created templates are shown respectively in Table 3, and Table 4.

The most important and also the most difficult case is Case 2, which we explained in section 2.1.

In this case, we use strings having 2 words of keyword to query. The following is the example.

Input: *A lot of trees were blown down in the recent storms.*

Keywords: *storm, blown.*

Table 5 shows the templates.

In this example, if there are less than n_l extracted adjectives, we have to use templates for only noun and only verb to query Google *N*-gram separately. We limited the number of list of adjectives because of the computational efficiency of adjective selection. In the proposed system, we set n_l = 20 based on the preliminary experiment. The result of this case will be the intersection of 2 queried adjectives lists.

2.3. Dataset

After having a list of adjectives, the problem is how to compute the similarity score between each adjective in the list and the keyword. To address this task, we need to create a group of words to query frequency based on Google *N*-gram. The method to create strings to query for keywords and each adjective is almost the same as in section 2.2. However, each adjective of created resulting list in section 2.2 is added to the template to query. The result is the frequency for each keyword and adjective. To illustrate this step, we use the case - sunrise is the keyword. The templates to retrieve the frequency of key noun sunrise and adjective in this case are presented in Table 6.

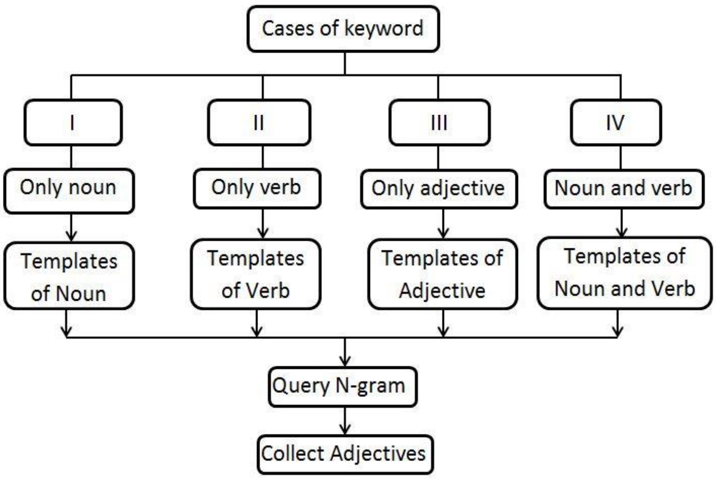

Figure 3: Overview of Adjective Collection

Table 1: Templates to query adjectives in the case there is only one noun

Templates	**Examples**
as + noun	*as sunrise*
Singular noun + to be	*sunrise is*
Capitalized singular noun + to be	*Sunrise is*
Plural noun + to be	*sunrises are*
Capitalized plural noun + to be	*Sunrises are*

2.4. Similarity measure

According to the analysis in section 2.1, we have cases of keywords: Case 1 - a single word and Case 2 - a pair of words. Therefore, we need 2 cases of measurements. There are:

- Two variants association measures are used for Case 1.
- Multi variants association measures are used for Case 2.

Table 2: Adjective list of *"sunrise"* query

Input	*Sunrise*
Output	*'beautiful', 'drier', 'certain', 'sharp', 'geological',…*

Table 3: Templates to query adjectives in the case there is only one verb

Templates of tenses of verb	**Examples**
Present	*Rain*
Present and 3rd singular	*Rains*
Past	*Rained*
Present continuous	*Raining*
Present perfect	*Rained*
Capitalize	*Raining*

Table 4: Templates to query adjectives in the case there is only one adjective

Templates	**Examples**
Adjective + as	*rainy as*
Adjective + and	*rainy and*
Query as the case that verb keeps the role of the keyword	(showed in table 1)
Query as the case that noun keeps the role of the keyword	(showed in table 3)

Table 5: Templates to query adjectives in the case keyword is a pair of noun and verb

Templates	**Examples**
Singular noun + present tense of verb	*storm blows*
Singular noun + past tense of verb	*storm blew*
Singular noun + past perfect tense of verb	*storm blown*
Singular noun + V-ing	*storm blowing*
Plural noun + present tense of verb	*storms blow*
Plural noun + past tense of verb	*storms blew*
Plural noun + past perfect tense of verb	*storms blown*
Plural noun + V-ing	*storms blowing*
Capitalized plural noun + present tense of verb	*Storms blow*
Capitalized plural noun + past tense of verb	*Storms blew*
Capitalized plural noun + past perfect tense of verb	*Storms blown*
Capitalized plural noun + V-ing	*Storms blowing*
Use verb and noun to query separately in the case there are not enough n_l adjectives	Showed in Table 1 and Table 3

Table 6: Templates to query dataset in the case there is only one noun

Templates	Examples
Singular noun + adjective	*sunrise beautiful*
Plural noun + adjective	*sunrises beautiful*
Capitalized plural noun + adjective	*Sunrises beautiful*

2.4.1. Two variants association measures for Case 1

Here, we propose a new concept of association, so we modify four popular co-occurrence measures: Jaccard, Overlap, Dice, and Pointwise mutual information (PMI) to compute semantic similarity using Google *N*-gram. Before introducing association measures, notations are presented in the following Table 7. In this paper, parameters are noun, verb and adjective.

- Jaccard Coefficient: It is often used in information retrieval. The measure was originally designed for binary vectors. It divides the number of equal features with the number of features in general. The Jaccard coefficient measure for two words is computed as

$$Jaccard(w_1, w_2) = \frac{F(w_1, w_2)}{F(w_1) + F(w_2) - F(w_1, w_2)} \quad . \quad (1)$$

- Dice Coefficient: It is very similar to the Jaccard measure and is also often used in information retrieval. The definition is as follows:

$$Dice_1 = \frac{2F(w_1,\ w_2)}{F(w_1) + F(w_2)} \quad . \quad (2)$$

- Overlap Coefficient: The measure is defined:

$$Overlap(w_1, w_2) = \frac{F(w_1, w_2)}{\min(F(w_1), F(w_2))} . \quad (3)$$

- Pointwise mutual information (PMI): It computes how often a lexeme and a feature co-occurrence, compared with what would be expected if they were independent. This measure is computed as

$$\mathrm{PMI}(\mathrm{w}_1,\ \mathrm{w}_2) = \log_2 \frac{P(w_1, w_2)}{P(w_1)(w_2)} \quad . \quad (4)$$

The occurrence score is calculated using two variation association measures between keyword and each adjective in the adjectives list. The result of the example whose keyword is *sunrise* is illustrated in the Table 8.

2.4.2. Multi variants association measures

The association will be the association between noun, verb and adjective. (Tim, 2011) explored two possible generalizations of pointwise mutual information (PMI) for multi-way co-occurrences and Dice for 3 parameters.

- Interaction information: Interaction information is based on the notion of conditional mutual information. Conditional mutual information is the mutual information of two random variables conditioned on the third one. Interaction information can be equally be defined for nv > 2 variables.

$$SI_1(w_1,w_2,w_3)=\log_2 \frac{P(w_1,w_2)P(w_2,w_3)P(w_1,w_3)}{P(w_1)P(w_2)P(w_3)P(w_1,w_2,w_3)} \quad . \qquad (5)$$

- Total correlation: Total correlation quantifies the amount of information that is shared among the different random variables, and thus expresses how related a particular group of random variables are

$$SI_2(w_1,w_2,w_3)=\log_2 \frac{P(w_1,w_2,w_3)}{P(w_1)P(w_2)P(w_3)} \quad . \quad (6)$$

- Dice 2: The extension of original Dice measure.

$$Dice_2(w_1,w_2,w_3)=\frac{F(w_1,w_2,w_3)}{F(w_1)+F(w_2)+F(w_3)} \quad . (7)$$

For our experiments, these measures are applied into the proposed system to compute the association between pair of words (noun and verb) and adjective. We use Table 9 to show the result of the example that *storm*, and *blown* are keywords.

However, one important point needs to notice in this case. As we especially mentioned in the section 2.2, if there are less than n_l extracted adjectives, we need to query adjective for noun and verb separately and intersect 2 resulted lists to have the final result. Therefore, the method to calculate the similarity scores is the similarity measurement between each adjective and each word of keywords independently using method as section 2.4.1.

Table 7: Notation

Notations	Description
F(w1)	Frequency of word 1 in corpus
F(w2)	Frequency of word 2 in corpus
F(w3)	Frequency of word 3 in corpus
Nw	Total number of words in corpus
F(w1,w2)	$\sum co-occurrence\ frequency\ of\ word\ 1\ and\ word\ 2$
F(w1,w3)	$\sum co-occurrence\ frequency\ of\ word\ 1\ and\ word\ 3$
F(w2,w3)	$\sum co-occurrence\ frequency\ of\ word\ 2\ and\ word\ 3$
F(w1,w2,w3)	$\sum co-occurrence\ frequency\ of\ word\ 1\ and\ word\ 2\ ,\ and\ word\ 3$
P(w1)	$F(w_1)/N_w$
P(w2)	$F(w_2)/N_w$
P(w3)	$F(w_3)/N_w$
P(w1,w2)	$F(w_1,w_2)/N_w$
P(w1,w3)	$F(w_1,w_3)/N_w$
P(w2,w3)	$F(w_2,w_3)/N_w$
P(w1,w2,w3)	$F(w_1,w_2,w_3)/N_w$

Table 8: Adjectives list of *"sunrise"* query

Adjective	PMI	Jaccard	Overlap	$Dice_1$
beautiful	15.964	0.001	0.003	0.002
Drier	14.266	0.0002	0.0003	0.001
Certain	8.207	0.0001	0.0002	0.0001
Best	6.964	0.00002	0.0005	0.0037
…..	…...	……	……	……

Table 9: Adjectives list of *"storm blown"* query

Adjective	SI_1	SI_2	$Dice_2$
Cute	1.0079	0.0003	0.0001
Atmosphere	9.0889	8.2228	0.0006
Evolutionary	4.2076	4.4170	0.0001
ceramic	3.1546	4.4365	0.0002
…..	…...	……	……

2.5. Sentiment:

In our research, we use *patterns.en* library to calculate and check the polarity orientation of the input and output.

For example:

Input: *A lot of trees were blown down in the recent storms.*

- Sentiment of input:
  sentiment("A lot of trees were blown down in the recent storms") = (-0.0556, 0.2639)
- Sentiment of its list of adjective after extracting:
  sentiment (*cute*) = (0.5, 1.0)

  sentiment(*atmosphere*) = (0, 0)

  sentiment (*severe*) = (-0.25, 0.25)

At this step, the sentiment of all adjectives is checked. All adjectives have the same polarity as input are kept and ordered for the next step ranking.

2.6. Adjectives Ranking

We again need to consider separate ranking procedure for the cases: single keyword and pair of words keyword.

2.6.1. Ranking method using for Case 1 (Section 2.1)

The method we use to rank is the average ranking. Depending on the situation, the vector of each adjective has 3 or 4 elements corresponding with the association measures (SI1, SI2, Dice) or (PMI, Jaccard, Overlap, Dice).

However, since association measures have different ranges, normalization scheme is thus necessary. In this application, the scale from 0 to 1 is used. It helps parameters have the same scale for a fair comparison between them. The normalized value is computed as following:

Table 10: Normalization Notation

Notation	Explanation
E	A column needs to normalize
E_{max}	Maximum value of column E
E_{min}	Minimum value of column E
e_i	Value for variable E in the ith row
v_n	Normalized value

Table 11: After normalization

Adjective	SI1	SI2	Dice2
Cute	0.007	0.008	0.007
Atmosphere	0.046	0.050	0.050
Evolutionary	0.017	0.014	0.010
Ceramic	0.003	0.004	0.003
…..	…...	……	……

$$If\,(E_{\max} = E_{\min}) \quad then\ v_n = 0.5$$

$$else\ \ v_n = \frac{e_i - E_{\min}}{E_{\max} - E_{\min}} \quad (8)$$

where the notations for normalization computation are shown in Table 10.

- Apply into our application:

List N_a adjectives, each column is value of each association measure. Scores will be normalized to let elements of each adjective have a fair role in ranking. Table 11 shows the result after normalization.

- Ranking

The way to rank is the usage of the averaged rank. It means we compute the average score of semantic association score of each adjective.

We calculate the average score for each adjective i of list: $r_i = \frac{\sum_j^m r_{ij}}{m}$, where m is the number of the association measures we used (in this case m = 4). The final ranking is obtained by ordering the average ranks.

The output of this task is the top n_a words having the highest average score and the same orientation with the input polarity orientation.

2.6.2. Ranking method in Case 2 (section 2.1)

We use the same way as section 2.6.1 to order adjectives. However, this case has special exception which is noticed in the section 2.4.2. The similarity score of noun, verb and adjectives are calculated separately noun - adjectives and verb - adjectives. Therefore, to rank adjectives, we apply the method in section 2.6.1 to each word of keyword and get two rankings, r_1 and r_2. Then, results need to be merged. The computation of ranking combination is proposed

$$r(x)=\frac{\frac{1}{r_1(x)}+\frac{1}{r_2(x)}}{1+\left|r_1(x)-r_{12}(x)\right|}\quad , \quad (9)$$

where: $r\ (x)$: rank of item x in list after combination two lists.

$r_1(x)$: rank of item x in list when list is affected by factor 1.

$r_2(x)$: rank of item x in list when list is affected by factor 2.

This proposed ranking method is used in the case we have a list of items and 2 factors affecting the order of list. The list will be ordered based on the factor in which the list interacts with. The goal of our method is combine two lists order together to optimize the order list. Finally, we select the top N_a adjectives in the ordered list whose polarity orientation is similar to the input polarity orientation.

3. EXPERIMENT AND EVALUATION

3.1. Experiment

In order to evaluate the effectiveness of the proposed system, we conducted an experiment. With each input, the results are lists of 3 adjectives, their order and their scores. Table 12 shows the results of the experiment.

Table 12: Experimental Result

N	Sentence	Adj1	S	Adj2	S	Adj3	S	Adj4	S	Adj5	S
1	It is snowy	White	0.72	Light	0.49	Cold	0.43	Solid	0.38	Scenic	0.37
2	It is windy	Vigorous	0.29	Breeze	0.27	Various	0.27	Accidental	0.27	Medical	0.27
3	It is cloudy	Momentary	0.62	White	0.52	Light	0.52	Black	0.50	Tag	0.45
4	It is rainy	Light	0.54	Sky	0.40	Heavy	0.34	Patriotic	0.32	Raindall	0.31
5	It will be sunny	Early	0.70	Real	0.63	Good	0.42	Competitive	0.37	Second	0.36
6	It is raining outside	Light	0.55	Sky	0.40	Heavy	0.38	Patriotic	0.32	Raindall	0.31
7	It is snowing outside	Light	0.62	White	0.56	Heavy	0.46	Solid	0.3	Regional	0.28
8	Sunrise	East	0.50	Sky	0.30	Bright	0.32	Beautiful	0.28	Geological	0.27
9	Sunset	Red	0.43	Ocean	0.42	Western	0.38	Super	0.36	Beautiful	0.30
10	Some people are going under the snow	Slow	0.53	First	0.49	White	0.46	Heavy	0.45	Central	0.43
11	An earthquake destroyed	Massive	0.43	Recent	0.43	Emergency	0.39	Powerful	0.32	Huge	0.31

	much of San Francisco.										
12	A lot of trees were blown down in the recent storms	Magical	0.31	Winter	0.27	Terrible	0.25	Noxious	0.25	Severe	0.25
13	Storm hits Oklahoma	Early	0.38	Real	0.35	Good	0.33	Competitive	0.32	Second	0.29
14	Summer comes	Early	0.63	Hard	0.56	Hot	0.54	Full	0.50	Expensive	0.49
15	Winter comes	Harsh	0.38	New	0.31	Beautiful	0.30	Long	0.29	Cold	0.27
16	Autumn comes	Alive	0.14	Love	0.14	Brown	0.11	Accessible	0.11	Cold	0.10
17	Spring comes	Hot	0.59	Important	0.57	New	0.56	Critical	0.53	Great	0.51
18	Rainbow appears	Beautiful	0.56	Southeast	0.47	Remarkable	0.46	Global	0.43	Yellow	0.34
19	Ice is melting	Arctic	0.61	Faster	0.48	Global	0.46	Antarctic	0.45	Northern	0.43
20	Outside the snow began to fall	Heavy	0.51	Fresh	0.42	White	0.40	Light	0.39	More	0.38

3.2. Evaluation Method

3.2.1. Method

Subjective experiment for the evaluation was carried out using 60 subjects (39 males and 21 females, 44 non-native speakers and 16 native speakers). The ages of subjects are older than 10 years old. We created and distributed the survey to let people judge output of the proposed system corresponding with the given input.

3.2.2. Materials

This survey consists of 20 concept sets, each set consisted of a sentence such as *"Storm hits Oklahoma"* and 5 adjectives that system conducted to express impression of the given sentence. To evaluate the proposed system imple-mentation is, we tried 4 types of different sentences:

- Sentences whose keyword is only one noun.
- Sentences whose keyword is an adjective.
- Sentences whose keyword is a noun.
- Sentences whose keywords are a noun and a verb.

The topic the proposed system treated is related to the weather, or the natural phenomena. We choose these topics be-cause they are commonly used, can be representative topics as given examples and are able to help us evaluate system easily and more accurately.

The grade scale is from 1 to 5. In each concept, grades are given for individual evaluation. The majority of work for subjects is that they would give the grade that best matches with how agreeable they thought about the outputs for the given input sentences. Table 13 summarizes the number of sentences in each group.

3.2.3. Procedure

After accessing the consent forms, the sub-jects read instructions from the computer screen. There are two parts of the survey that the users need to finish. The first part asks personal information such as gender, age, nationality, and level of English language. The main part is the second part. In this part, the subjects need to indicate their level of agreement with generated adjective corresponding with the given input in each concept on scale from 1 to 5.

3.3. Evaluation Result

Table 14 shows summary statistics for the 4 samples of data. As the table shows the score participants marked stretch from 1 to 5. The common evaluation is agreement which we can see from the average grade, at 3.1. To compare the difference grading between groups of sentences, Fig.4 shows the frequency of subjects for each type of sentence in each grade. Overall, it is clear that noun-group has higher score than the other groups and satisfied with the majority of people (more than 40 people graded 4 and 5). For other groups, the popular grade is 3.0. To be clearer, Fig.5 shows how different is it in the grades between groups. Obviously, the efficiency of noun group compared to other group is much higher.

Table.15 and Table.16 show which means of groups are significantly different from which others. Most of the pairs are significant difference, but just only one pair adjective and verb is not and the grades for these pairs are not that high. It means that adjective-group and verb-group are considered as two non-effective working groups.

4. CONCLUSION

A new system of the impression estimation of a short sentence has been presented in this paper. In order to obtain keywords, we analyze and pre-processed the input. They are then used to make up templates to collect adjectives from queried data using Google *N*-gram. The co-occurrence scores between keywords and adjectives are computed based on semantic similarity computational measurements: Jaccard coefficient, Dice coefficient, Overlap coefficient, and Pointwise mutual information. After that, polarities between input and adjectives are checked using the sentiment library *patterns.en*. Finally, the adjectives are ranked and top n_a adjectives are considered as an output of system. For example, the experiments were carried out and got fairly good result. With the input "*it is snowy*", the results are *white* (0.70), *light* (0.49), *cold* (0.43), *solid* (0.38), and *scenic* (0.37)

In our future work, we will improve more in the tasks of keyword extraction and semantic similarity methods to make the proposed system working well with complex inputs.

Table 13: Number of sentences

Type	Number
Adjective	5
Verb	2
Noun	2
Verb and noun	11

Table 14: Experiment average result

Max	Min	Mean	Median	Stdev
5	1	3.1	3	1.2

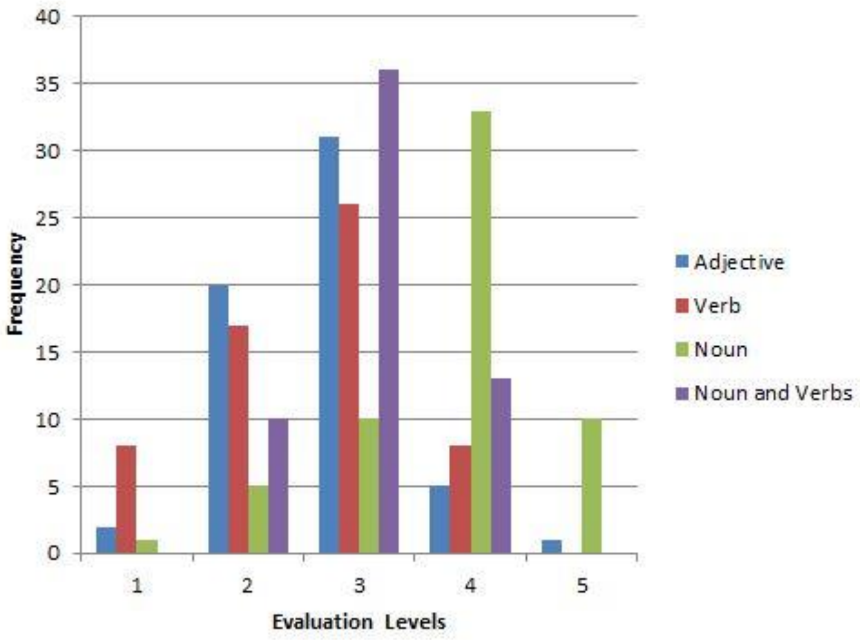

Figure 4: Evaluation levels

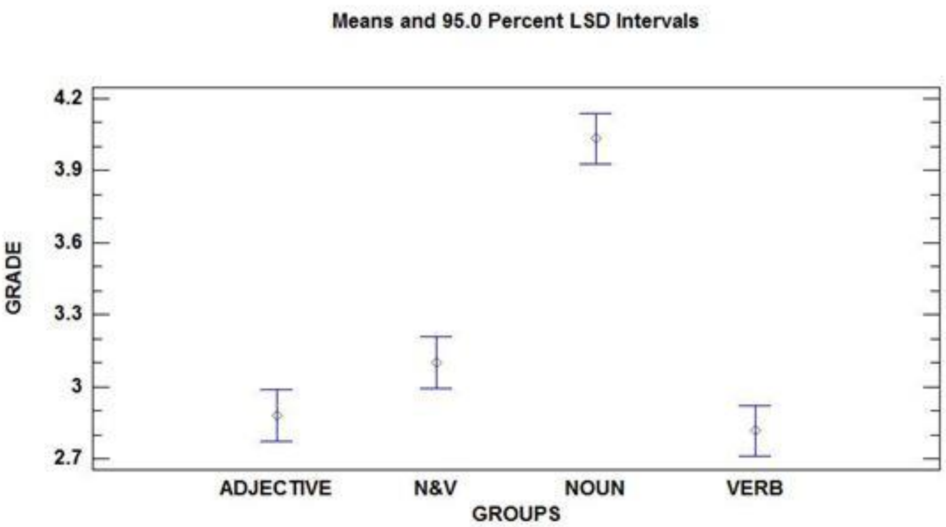

Figure 5: Group Comparison

Table 15: Comparison result

Groups	Count	LS Mean
Verb	60	2.82
Adjective	60	2.88
Noun and verb	60	3.10
Noun	60	4.03

Table 16: Difference between groups

Contrast	Sig.	Difference	+/- Limits
Adjective – Noun and Verb	*	-0.223	0.213808
Adjective – Noun	*	-1.15417	0.213808
Adjective - Verb		0.0625	0.213808
Noun and verb – Noun	*	-0.931167	0.213808
Noun and verb – Verb	*	0.2855	0.213808
Noun – Verb	*	1.21667	0.213808

REFERENCES

Linlin Li, Benjamin Roth, and Caroline Sporleder (2010). Topic models for word sense disambiguation and token-based idiom detection. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, (pp. 1138–1147), Uppsala, Sweden: ACL.

Saini, M., & Sharma, D., & Gupta, P.K. (2011). Enhancing information retrieval efficiency using semantic-based-combined-similarity-measure. Image Information Processing (ICIIP), International Conference on (pp. 1-4), Waknaghat, India.

Prodromos Malakasiotis (2009). Paraphrase Recognition Using Machine Learning to Combine Similarity Measures. Proceedings of the ACL-IJCNLP 2009 Student Research Workshop (pp. 27-35), Suntec, Singapore.

Ramiz Aliguliyev (2009). A new sentence similarity measure and sentence based extractive technique for automatic text summarization. Expert Systems with Applications, Vol.36, pp.7764–7772.

Jian Hu, Lujun Fang, Yang Cao, Hua-Jun Zeng, Hua Li, Qiang Yang, and Zheng Chen (2008). Enhancing text clustering by leveraging Wikipedia semantics. Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval (pp. 179-186). Pisa, Italy.

Alexander Budanitsky and Graeme Hirst (2006). Evaluating WordNet-based Measures of Lexical Semantic Relatedness. Computational Linguistics, Vol.32, num.1, pp. 13-47.

Mehran Sahami, Timothy Heilman (2006). A web-based kernel function for measuring the similarity of short text snippets. WWW '06 Proceedings of the 15th international conference on World Wide Web, Vol. 2, (pp. 377 – 386), Edinburgh, Scotland.

Huirong Yang, Pengbin Fu, Baocai Yin, Mengduo Ma, and Yanyan Tang (2011). Computer Software and Applications Conference (COMPSAC) 2011 IEEE 35th Annual, Vol. 69, Munich, Germany.

Danushka Bollegala, Yutaka Matsuo, and Mitsuru Ishizuka (2011). A Web Search Engine-based Approach to Measure Semantic Similarity between Words. Knowledge and Data Engineering, IEEE Transactions on, Vol. 23 (pp.977 – 990).

Michael Strube and Simone Paolo Ponzetto (2006). WikiRelate! computing semantic relatedness using Wikipedia. Proceedings of the 21st national conference on Artificial intelligence - Vol.2 (pp.1419 – 1424). Boston, USA.

Lun-wei Ku and Yong-sheng Lo and Hsin-hsi Chen (2007). Using Polarity Scores of Words for Sentence-level Opinion Extraction. Proceedings of NTCIR-6 Workshop Meeting, May 15-18, Tokyo, Japan, 2007.

Subrahmanian V. S. and Reforgiato Diego (2008). AVA: Adjective-Verb-Adverb Combinations for Sentiment Analysis", IEEE Intelligent Systems, Vol.23 (pp. 43-50).

Samaneh Moghaddam and Fred Popowich (2010). Opinion Polarity Identifi-cation through Adjectives. CoRR, abs/1011.4623, 2010.

Tim Van de Cruys (2011). Two Multivariate Generalizations of Pointwise Mutual Information",Proceedings of the Workshop on Distributional Semantics and Compositionality (DiSCo2011), (pp. 26-20).

Aminul Islam and Evangelos Milios and V Keelj (2012). Comparing Word Re-latedness Measures Based on Google n-grams. Proceedings of COLING 2012: Posters (pp.495-506). Mumbai, India.

Kerstin Denecke (2008). Using SentiWordNet for multilingual sentiment analysis. Data Engineering Workshop, ICDEW 2008. IEEE 24th International Conference on, Vol. 72 (pp.507 – 512), Cancún, Mexico.

Hugo Liu , Push Singh (2004). ConceptNet: A Practical Commonsense Rea-soning Toolkit", BT TECHNOLOGY JOURNAL, Vol.22 (pp. 211-226).

Rudi Cilibrasi, Paul Vitanyi (2007). The Google Similarity Distance. IEEE Trans. on Knowl. and Data Eng., Vol. 19 (pp. 370-383).

Rosalind Picard and Jonathan Klein (2002, ScienceDirect). Computers that recognise and respond to user emotion: theoretical and practical implications. Interacting with computers, Vol. 14 (pp.1-20).

Danushka Bollegala, Yutaka Matsuo and Mitsuru Ishizuka (2004). Keyword Extraction From A Single Document Using Word Co-Occurrence Statistical Information", International Journal on Artificial Intelligence Tools, Vol. 13.

Ulli Waltinger (2009). Polarity reinforcement: Sentiment polarity identification by means of social semantics", AFRICON, 2009 (pp. 1-6).

Kazuyuki Matsumoto, Junko Minato, Fuji Ren, Shingo Kuroiwa (2005, IEEE). Estimating Human Emotions Using Wording and Sentence Patterns. International Conference on Information Acquisition, (pp. 422-426).

Rosalind Picard (2003). Affective computing: Challenges. International Journal of HumanComputer Studies, Vol.59.

Schumaker, Robert P. and Liu, Ying and Ginsburg, Mark and Chen, Hsinchun (2006). Evaluating mass knowledge acquisition using the ALICE chatterbot: the AZ-ALICE dialog system. Int. J. Hum.-Comput. Stud., Vol.64 (pp.1132-1140).

Clips, computational linguistics and psycholinguistics research center, http://www.clips.ua.ac.be/pages/pattern-en.

BIOGRAPHY

Short biography of author 1 (100 words max) (leave blank for reviewing)

Nguyen Thi Thu An graduated with BA in Computer Science from Can Tho University, Can Tho, Viet Nam in 2008. After graduation, she worked as a lecturer at College of Information and Communication Technology of that university. She returned to school on Mext Scholarship. She received her Msc from Keio University, Yokohama, Japan (2013) where she is a first doctoral student now. Her research interests focus on natural language processing and AI.

Short biography of author 2 (100 words max) (leave blank for reviewing)

Masafumi Hagiwara received his B.E., M.E. and Ph.D degrees in electrical engineering from Keio University, Yokohama, Japan, in 1982, 1984 and 1987, respectively. Since 1987 he has been with Keio University, where he is now a Professor. From 1991 to 1993, he was a visiting scholar at Stanford University. He received IEEE Consumer Electronics Society Chester Sall Award, Author Award from the Japan Society of Fuzzy Theory and Systems, Technical Award and Paper Award from Japan Society of Kansei Engineering, best research award from Japanese Neural Network Society in 1990, 1996, 2003, 2004, and 2013, respectively.