

Generation of Image by Sentence Based on Impression Words in Image

Shohei Namisato^{1*}, Kazuaki Yokota¹, Hiroki Yamaoka¹, Sho Ooi², Mutsuo Sano³

¹ Graduate School of Information Science and Technology, Graduate School of Osaka Institute of Technology, Osaka Japan.

² Faculty of Information Science and Engineering, Ritsumeikan University, Shiga, Japan.

³ Faculty of Information Science and Technology, Osaka Institute of Technology, Osaka Japan.

* Corresponding author. Tel.: +81(72) 866-5301; email: m1m18a18@st.oit.ac.jp

Manuscript submitted March 11, 2019; accepted May 10, 2019.

doi: 10.17706/jcp.14.5.349-358

Abstract: The use of social network services (SNS) have become widespread in recent years, and it is used to express the situation of individual users by using short sentences and images. Especially, young people prefer to express themselves by using short sentences on Twitter and pictures on Instagram. However, CAMPUS LIFE DATA 2017 suggested that young people do few read a long sentence in a newspaper, journal, and more. We think that it is important for people who have not read long sentences to express long sentences with images. Therefore, we aim to express images and short sentences in SNSs from a long sentence in a newspaper, journal, and other reading material in this research. Specifically, we extract the impression of a person looking at an image or after reading a sentence. We calculate the similarity of each extracted word and generate a new image similar to the impression of the sentence. The results indicate that the system could estimate the impression of human beings. However, it was found that a large number of impression words and images are required to improve accuracy.

Key words: Impression words, image processing, morphological analysis, text to image.

1. Introduction

The use of social network services (e.g, Facebook [1], Twitter [2], and Instagram [3]) has become widespread in recent years, and individuals express themselves through these platforms by using short sentences and images. Especially, young people prefer to express themselves by using short sentences on Twitter and pictures on Instagram. However, CAMPUS LIFE DATA 2017 suggested that young people do few read a long sentence in a newspaper, journal, and more [4]. We think that it is important for people who have not read long sentences to express long sentences with images. Therefore, we considered that it is important to convert long sentences into images as shown in Fig. 1. By expressing sentences in images, young people who do not a good at reading sentences can understand the contents of the sentence. In order to generate an image from a sentence, there are the following problems.

- 1) Associate an image of a sentence with an image.
- 2) Combine multiple images.
- 3) Understand sentences and make stories using existing images.
- 4) Understand sentences and generate new images.

In a research using impression words, Kurita *et al.* created a database of paintings by using impression

words and constructed a system to search by calculating the similarity of impression words [5]. Kumamoto *et al.* used a method of searching for something that meets judgment criteria such as an individual's preference and sensitivity in large digital contents [6]. Yohida *et al.* used impression words to create a database of pictures in ART MUSEUM [7]. Hayashi *et al.* proposed a method to estimate impression words by using an image search system and neural network and automatically extracted images corresponding to impression words [8]. Yamamoto *et al.* visualized the relationship between impression words and values in individuals [9].

In another study, Takagi *et al.* proposed a method for quantitatively evaluating the difference in the nuances of impression words expressing emotions and aesthetic feelings [10], Akai *et al.* proposed an analysis method based on impression words. Therefore, impression words are used for constructing a database and evaluate sensitivity.

In recent years, methods using deep learning are popular as a method of generating images from texts. Tao *et al.* [11] proposed Attentional Generative Adversarial Network (AttnGAN) that can synthesize fine-grained details at different subregions of the image by paying attentions to the relevant words in the natural language description. Han *et al.* [12] proposed Stacked Generative Adversarial Networks (StackGAN) to generate 256×256 photo-realistic images conditioned on text descriptions. Hasegawa *et al.* [13] proposed the automatic system for comic creation from a TV script. Namiki *et al.* [14] proposed an automatic generation system of a piece of the comic from a newspaper article.

We aim to express images and short sentences in SNSs from a long sentence in a newspaper, journal, and other reading material in this research. Specifically, we extract the impression of a person looking at the image and the impression seen from the sentence as shown in problem 1). We calculate the similarity of each extracted impression word and generate a new image similar to the impression of the sentence. In addition, an impression word is a word that expresses the direct feeling that is perceived by an observer after seeing a word. An impression word is used to strongly express something that cannot be forgotten [15].

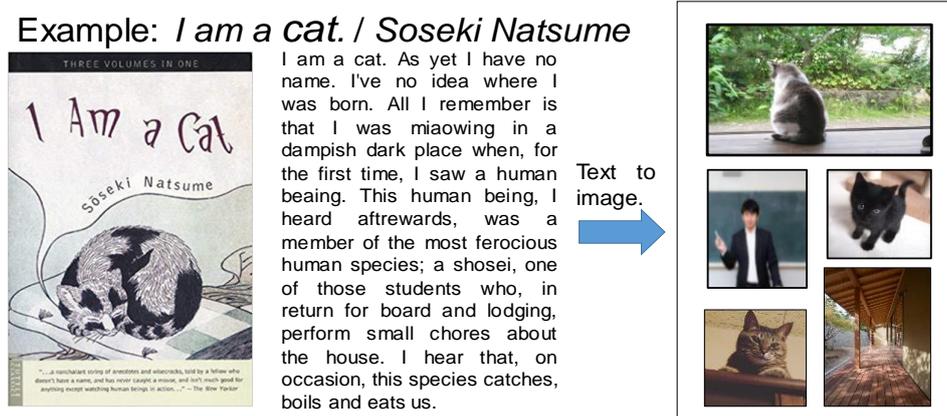


Fig. 1. Example of text to image.

2. Method to Create Image from Impression Words

The flowchart of generating an image from impression words is shown in Fig. 2. The method proposed by this research consists of the following steps 1 to 4.

Step.1 Extract impression words from image

To extract the impression word from the image, a questionnaire survey is conducted. An impression word histogram of the image is created from the results of the survey.

Step.2 Extract impression words from sentence

We extract impression words from the results of the morphological analysis of sentences and create impression word histograms of sentences.

Step.3 Calculate similar of impression words histogram

We calculate the similarity by using the cosine similarity to the impression word histogram of images and sentences obtained in Step 1 and Step 2.

Step.4 Create new similar image

Finally, we obtain N original images showing large values of similarity obtained in Step 3 and generate a composite image from N original images.

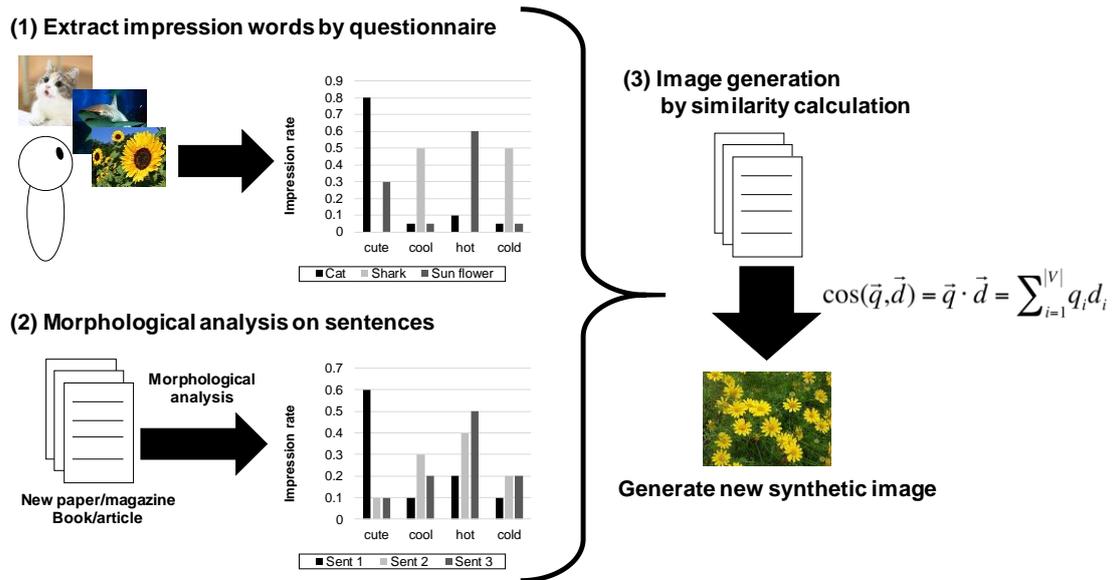


Fig. 2. Method of image generating based on impression words.

2.1. Extract Impression Words from Image

We conducted a questionnaire survey on 15 experimental cooperators to extract impression words from images. The number of images was set to 100 with 20 images for each category. "Category," "food," "animal," "plant," and "color" were the five categories. In addition, we developed 50 composite images based on 100 images. Here, we make composite images based on the order of high questionnaire result. Ten types of impression words were used: "cool," "cute," "hot," "cold," "soft," "hard," "sieve," "new," "easy," and "difficult." The participants answered using a selection formula. An example of each category image is shown in Fig. 3, an example of a synthesized image is shown in Fig. 4, and a questionnaire conducted by the questionnaire is shown in Fig. 5.

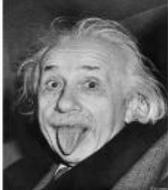
				
Historical Person	Plant	Color	Foods	Animals

Fig. 3. Example of image each category.



Fig. 4. Example of combination image.

questionnaire

* Required

What kind of impression do you see by seeing the picture below? Please select one or more from the choices.

Q1 *



New
 Old

Fig. 5. Example of a questionnaire.

We obtain an impression word histogram H_i based on a result of their questionnaires. As example of such impression word histograms, Plant in Fig. 3 (sunflower) is {"cool," "cute," "hot," "cold," "soft," "hard," "sieve," "new," "easy," "difficult."} = {0.15, 0.22, 0.27, 0.00, 0.10, 0.02, 0.00, 0.10, 0.15, 0.00}.

2.2. Extract Impression Words from Sentence

We conducted a morphological analysis on long sentences such as those found in newspaper and magazine articles by using Janome [16], a morphological analysis library. We used mecab-ipadic-2.7.0-20070801 for Janome's included dictionary in this study. From the results of the morphological analysis, the number of occurrences of each stamp word similar to the image from among the ten impression words is counted, and the frequency of occurrence of each stamp word is considered as the impression word histogram of the sentence. However, there were cases where impression words defined by sentences do not appear if only the defined impression words are used. Thus, the method of selecting synonyms was the same as that used in the study of Kumamoto *et al.* [5], the similarity of the defined impression words treat as words.

2.3. Calculate Similar of Impression Words Histogram

When a certain impression word histogram is given by H_1, H_2 , the cos similarity is expressed by equation (1). Here, the value that the cos similarity can take is given by equation (2). As the value of cos similarity

approaches 1, it can be determined that the degree of similarity between the two objects is large, and as the value approaches 0, it can be judged that the degree of similarity is small.

$$\widehat{cos}(H_1, H_2) = \frac{H_1 \cdot H_2}{|H_1| |H_2|} \quad (1)$$

$$0 \leq \widehat{cos}(H_1, H_2) \leq 1 \quad (2)$$

When, H_1 is the impression word histogram of image and H_2 is the impression word histogram of sentence.

2.4. Create New Similar Image

An image having an impression word with a high degree of similarity is calculated from the similarity calculation result described in Section 2.3 is displayed. In this study, we prepare a composite image beforehand and conduct a questionnaire on the impression of the composite image, calculate the similarity between the impression word histogram of the sentence, print the histogram of the composite image, and display the largest synthetic image. Here, when the number of sentences is N , the result is shown in Expression (3).

$$result = \max_{1, \dots, N} \widehat{cos}(H_1, H_2) \quad (3)$$

In this research, we present the result that is a top 5 high degree of similarity in learning images.

3. Experiments

For the experiment, we instructed the experiment collaborators to read the target sentences and administered a questionnaire to ascertain their impressions. We extracted impression words from the questionnaire. From the extracted results, we determined the impression word histogram in the sentence and calculated the similarity with the impression word histogram of the image collected from the questionnaire in advance. As a result of similarity calculation, it is assumed that the image of the top 5 is outputted and is an impression image obtained from the sentence. Next, to ascertain whether the image outputted by calculation is human emotion or not, 0 images selected from sentences are selected from the image group of 5 categories, who have had questionnaires taken from the sentences in advance. The degree of matching between the image obtained by the calculation and the image obtained by human sensibility is taken as the performance of the system.

We conducted experiments with 15 students. The sentences we used are as follows;

Sentence 1 "Kokoro, Soseki Natsume" [17].

Sentence 2 "Kyoto News Web, News reports" [18].

Sentence 3 "Kyoto News Web, News reports" [19].

Sentence 4 "News Paper, Yomiuri News Paper" [20].

Sentence 5 "Light Novel, The Boundary of Emptiness" [21].

4. Result and Discussion

First, Table 1 shows the top two positions of impression word histograms obtained from sentences 1 to 5.

Sentence 1 is Kokoro by written Soseki Natsume. This sentence uses words that are not used much in modern times. For example, its word is "telegram" and more. As a result, we think that the system chose "Old" as the impression words. Moreover, the reason why the proportion of "hard" is the second reason is because it seems that it is a hard sentence from an old word used in a sentence.

Since sentence 2 conveys a hot summer day, there are many words describing heat, such as "fierce heat" and "heat stroke," in the sentences. Therefore, "Atsui" is considered to be the highest proportion of impression words.

Table 1. A Result of Impression Words and Rate

Sentence number	First impression words	First impression words rate (%)	Second impression words	Second impression words rate (%)
Sent. 1	Old	46.7	Hard	26.7
Sent. 2	Hot	53.3	Hard	26.7
Sent. 3	Difficult / New	26.7	New / Simple	20.0
Sent. 4	New	100	None	-
Sent. 5	Cool	100	None	-

Sentence 3 is a news article that conveys the contents of Arashiyama, which is recovering from the damage from the western Japan torrential rainfall.

There is no phenomenon wherein sentence 3 concentrates on one impression word, like in sentence 1 or sentence 2, and as a factor, it can be made to correspond to any of the 10 impression words defined in this research. Such a word does not appear, and it seems that it became the result wherein individual differences of experiment collaborators appeared.

Next, an image in which a human reads a sentence and images is shown in Fig. 6. The result of sentence 1, from is a lot of fish and blue images, is shown in Fig. 6. The results when the experiment participants read the sentences were "old" and "hard"; however, when selecting an image, the image of "cold" impression was chosen. In sentence 2, many images having the same "hot" as the impression words of the text were selected as impression words. In sentence 3 is written the words that are the words "river", "heavy rain" and more. Therefore, we think that a brown image as the impression of "earth and sand", a blue image as the impression of "rain (water)" and "fish" are selected. In sentence 4, I think that an image that reminded me of money was chosen because it is an article about money. Sentence 5 is only the beginning of the book, giving a frightening impression when reading only this part. Therefore, I think that images such as dark colors, sharks, and crocodiles were chosen.

Then, an image representing the sentences obtained by the calculation results is shown in Fig. 7. As a result of Fig. 7, sentence 1 outputted many blue images in this system. We think that sentence 1 was written because synonyms of "strange" and "cold" were written. As sentence 2 was written with the words "hard" and "strong", we think that an image with a "hard" impression was chosen. In sentence 3, because the words "new" and "calm" were in the sentence, we think that an image with the impression of "new" or "easy" was chosen. Since sentences 4 and 5 did not include the defined impression, the system could not create an impression word histogram of sentences. As a result of these studies, we think that it is necessary to increase the impression words to be defined, as future research.

Finally, the degree of agreement between the image selected by humans and the image obtained by calculation is found, and the results are shown in Fig. 8. It can be seen from Fig. 8 that there was not much agreement between the image selected by a person and the image obtained by calculation. I think that this is because there were few impression words defined in this study.

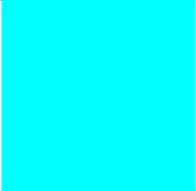
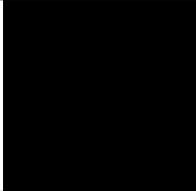
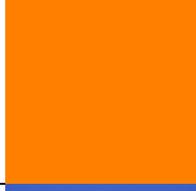
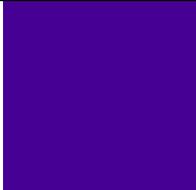
< Sent. 1 >	< Sent. 2 >	< Sent. 3 >	< Sent. 4 >	< Sent. 5 >
				
				
				
				
				

Fig. 6. A set of images felt by a human reading sentences.

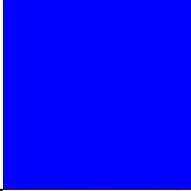
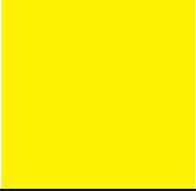
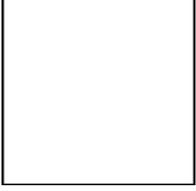
< Sent. 1 >	< Sent. 2 >	< Sent. 3 >	< Sent. 4 >	< Sent. 5 >
				
				
				



Fig. 7. A set of images calculated by system.

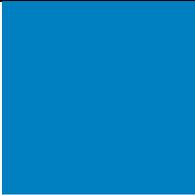
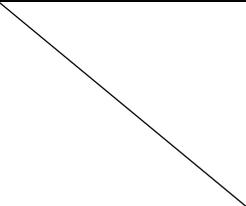
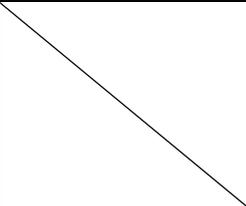
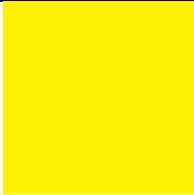
< Sent. 1 >	< Sent. 2 >	< Sent. 3 >	< Sent. 4 >	< Sent. 5 >
				

Fig. 8. Images that agree with calculation results and human results.

5. Conclusion

It was confirmed by the proposed method that the impression word of the sentence and the impression word of the composite image displayed from the sentence were the same. However, at the present stage, the impression word parameter closest to the composite image was calculated and displayed by composing the synthesized image in advance.

As a policy for future research, it is necessary to devise a method that selects images with high similarity between the impression word parameter of the image and the impression word parameter of the sentence, and to generate a composite image based on the selected images. In addition, it is thought that the main color, as an element that determines the impression of the image, is greatly involved, and many studies on the impression of color have been conducted. We will reflect these in future research.

References

- [1] Facebook, Inc.: "Facebook," Facebook, Inc.. Retrieved July 27, 2018, from <https://www.facebook.com/>
- [2] Twitter, Inc.: "twitter," Twitter, Inc.. Retrieved July 27, 2018, from <https://twitter.com/>
- [3] Instagram, LLC: "Instagram," Facebook. Retrieved July 27, 2018, from <https://www.instagram.com/?hl=ja>
- [4] National Federation of University Co-operative Associations (NFUCA): Campus life data 2017. Retrieved July 27, 2018, from <http://www.univcoop.or.jp/press/life/report.html>
- [5] Kumamoto, T., & Ohta, K. (2003). Selection of impression words for impression-based retrieval. *IPSJ Journal*, 44(7), 1808-1811.
- [6] KURITA, T., Kato, T., Fukuda, I., & Sakakura, A. (1992). Sense retrieval on a image database of full color paintings. *IPSJ Journal*, 33(11), 1373-1383.
- [7] Yoshida, K., Kato, T., & Yanaru, T. (1998). Image retrieval system using impression words. *Proceedings of 1998 IEEE International Conference on Systems* (pp. 2780-2784).

- [8] Hayashi, T., & Hagiwara M. (1997). An image retrieval system to estimate impression words from images using a neural network. *Proceedings of IEEE International Conference on Systems* (pp. 150-155).
- [9] Yamamoto, K., Kojima, T., Yoshikawa, T., & Furuhashi T. (2005). A basic study on discovering relationships of impression words among individuals using visualization method. *Proceedings of IEEE Workshop on Advanced Robotics and its Social Impacts* (pp. 206-209).
- [10] Taka, G. S., Isomoto, Y., & Nozaki, H. (2000). Quantitative estimation method for discrimination of nuance among impression words. *Proceedings of 2000 IEEE International Conference on Industrial Electronics, Control and Instrumentation* (pp. 1598-1603).
- [11] Xu, T., Zhang, P. C., Huang, Q. Y., Zhang, H., Gan, Z., Huang, X. L., & He, X. D. (2017). AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks. *Proceedings of 2017 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1316-1324).
- [12] Zhang, H., Xu, T., Li, H. S., Zhang, S. T., Wang, X. G., Huang, X. L., & Metaxas, D. (2017). StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks. *Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 5908-5916).
- [13] Akai, S., Hochin, T., & Nomiya, H. (2014). Analysis method based on impression words for impression evaluation method by space. *Proceedings of 2014 IIAI 3rd International Conference on Advanced Applied Informatics* (pp. 623-629).
- [14] Hasegawa, M., & Hayashi, M. (2002). Automatic generation of comic strip from TVML script. *The Journal of the Society for Art and Science*, 1(1), 15-21.
- [15] Namiki, K., & Hagiwara, M. (2014). An automatic generation system of the comic from a newspaper article. *Journal of Japan Society of Kansei Engineering*, 13(4), 501-509.
- [16] *Janome v0.3 Documentation (ja)*. Retrieved July 25, 2018, from <http://mocobeta.github.io/janome/>
- [17] Soseki, N. (2018). *Kokoro, Aozora Bunko*. Retrieved July 25, 2018, from https://www.aozora.gr.jp/cards/000148/files/773_14560.html
- [18] Kyoto news web: Severe vigilance against heat stroke following a hot summer day. Retrieved July 25, 2018, from <https://www3.nhk.or.jp/lnews/kyoto/20180724/2010001036.html>
- [19] Kyoto news web: Restarting heavy rain damage building newly built boat. Retrieved July 25, 2018, from <https://www3.nhk.or.jp/lnews/kyoto/20180725/2010001041.html>
- [20] *Yomiuri News Paper: Gift certificate of 25,000 yen*, 2018.
- [21] Kinoko, N. (2007). *The Boundary of Emptiness*. Dengeki Bunko.



Shohei Namisato received his B.S. degree in information science from Osaka Institute of Technology in 2018. Currently, he is a master student with Graduate School of Information Science and Technology in Osaka Institute of Technology, Japan. His research interests include cognitive science and virtual reality. He is a member of Information Processing Society of Japan (IPSJ).



Kazuaki Yokota received his B.S. degree in information science from Osaka Institute of Technology in 2018. Currently, he is a master student with Graduate School of Information Science and Technology in Osaka Institute of Technology, Japan. His research interests include cognitive science and human computer interaction and human robot interaction. He is a member of Information Processing Society of Japan (IPSJ).



Hiroki Yamaoka received his B.S. degree in information science from Osaka Institute of Technology in 2018. Currently, he is a master student with Graduate School of Information Science and Technology in Osaka Institute of Technology, Japan. His research interests include cognitive science and human computer interaction. He is a member of the Japanese Society for Artificial Intelligence (JSAI).



Sho Ooi received his B.S., M.S., and Ph.D. degree in information science from Osaka Institute of Technology in 2013, 2015, and 2018, respectively. Currently, he is an assistant professor with Faculty of Information Science and Engineering in Ritsumeikan University, Japan. His research interests include computer vision, cognitive science, and pattern recognition. He is a member of Information Processing Society of Japan (IPSJ), the Institute of Electronics, Information and Communication Engineers (IEICE), the Institute of Image Electronics Engineers of Japan (IEEEJ), and Robot Society of Japan (RSJ).



Mutsuo Sano received his B.S., M.S., and Ph.D. degree in engineering from Kyoto Institute of Technology in 1981, and Kyoto University in 1983, and 1995, respectively. Between April, 1983 and March, 1985. He was engaged in the research and robot vision, image processing and contents management of Nippon Telegraph and Telephone Corporation (NTT). Currently, he is a professor and a dean with Faculty of Information Science and Technology in Osaka Institute of Technology, Japan. His research interests include computer vision, cognitive science, cooking media, cooking support, and human robot communication. He is a member of Information Processing Society of Japan (IPSJ), the Institute of Electronics, Information and Communication Engineers (IEICE), the Institute of Image Electronics Engineers of Japan (IEEEJ), Robot Society of Japan (RSJ), the Japanese Society for Artificial Intelligence (JSAI), the Institute of Image Information and Television Engineers (ITE), and IEEE.