Catchcopy Generation for Fashion Goods using Visual Metaphor

Dan WANG†, Ryota MIBAYASHI††, and Hiroaki OHSHIMA†,††

† Graduate School of Information Science, University of Hyogo
7–1–28 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650–0047, Japan
†† Graduate School of Applied Informatics, University of Hyogo
7–1–28 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650–0047, Japan
E-mail: †ad21h036@gsis.u-hyogo.ac.jp, ††{aa20r511,ohshima}@ai.u-hyogo.ac.jp

Abstract In this study, we propose a method for generating catchcopies for fashion goods using visual metaphor. The input of the proposed method consists of a product image and its product description. For example, for a red jacket with good stretch, the catchcopy generated by the proposed method will be “leaping fire of youth”. For a pair of black high heels that can create an elegant temperament, the generated catchcopy will be “beautiful slender swallows”. The generated catchcopy contains a visual metaphor. A visual metaphor is something that is similar in appearance to a fashion goods and has a specific appearance. The catchcopy generation with a visual metaphor will bring inspiration to poster design or any other form of propaganda. The process of the proposed method consists of two parts. First, we build three deep learning models to predict the similarity between fashion goods and metaphor on color, shape, and texture. After that, we use the obtained visual metaphor and product description to generate catchcopy.

Key words Metaphor, Image Recognition, Image similarity

1. Introduction

For design of magazine ads and posters, there is an attractive way for creating a catchcopy with metaphor. The metaphor is something similar to the goods. Using this way, appeal of the goods can be conveyed to consumers through metaphors. For examples, a pair of white sneakers may thinked cool with “supple snow leopard”. A colorful bracelet with “little fruits candy” may looks more cuter. The use of metaphors in fashion goods promotion can not only highlight the appeal of fashion goods, but also bring consumers a sense of freshness. Forceville’s study summarized in detail the role played by visual metaphors in advertising [7]. Since advertising “has the explicit purpose of promoting a product or service, ” it is appropriate to test the theory of visual metaphor in advertisements. The persuasive effects of visual metaphors in advertising are also emphasized in early studies [10] [2].

The goal of this study is to automatically generate catchopies with visual metaphors based on product images and description of fashion goods. Figure 1 shows an example of input and output of this study. For a pair of black, curvaceous heels, we expect the generated catchcopy to be “Beautiful delicate swallow”. The word “swallow” in the generated catchcopy is defined as “metaphor word”. The

Figure 1 Example of input and output for this study

words “beautiful” and “delicate” appear together with the visual metaphor, are defined as “rhetoric part”. The words in rhetoric can be noun, verb, adverb or adjective, and the number of words is two. Modifiers exist in order to add a literary character to the catchcopy.

The innovation of this study lies in the retrieval of visual metaphors, i.e., finding something that look similar to the fashion goods in within a specific domain. The implementation of visual metaphor retrieval is not simply the same as searching for similar images. Visual metaphor retrieval based on fashion goods promotion requires not only judging the consistency of the appearance, but also defining what is a good metaphor. Once metaphors are defined, we can decide the scope of the search and ensure that the output metaphors are positive publicity. The definition of metaphors is a difficult point in this study.

This study focuses on the application of visual metaphors to generate catchcopies, which is an application of visual metaphor retrieval in advertising. However, visual metaphor retrieval can be applied not only to advertising, but also to a broader scope. For example, visual metaphors can be used in chat systems to generate humorous dialogues. Visual metaphors can be used in literature to inspire writers. In addition, the use of visual metaphors in education, games, etc. is also expected.

2. Related Work

There are researches focusing on visual metaphors in advertising. Forceville's study discussed two term of visual metaphors in advertising, literal and figurative, and the transfer of properties from figurative metaphors to literal metaphors [5]. In Philip's study, he describes nine different kinds of visual metaphors applied in advertising campaigns and their consumer reactions [15]. The examples he gives in paper emphasize certain characteristics of the goods through images containing metaphors. Forceville's research also gives various cases based on visual metaphors. One such example is the use of marshmallows to describe the style of clothing and its softness [6]. In Bulmer's study, he pointed out that visual metaphors have been used to very good effect in international advertising. Due to cultural and linguistic differences, Using textual advertising in an international environment brings limited promotional effect, while pictorial advertising can be simply understood by people in different countries [3]. The positive effect of visual metaphors on advertising has also been mentioned in many recent studies [4] [19].

In addition, researches on the automatic generation of advertising slogans has been very popular in recent years [8]. Alnajjar's research on generating sentences based on product features from keywords [1]. The system proposed in Özbal's research allows user to specify the number of words to be generated. The system can automatically generate rhyming phrases based on color and category of the goods [13]. Munigala's proposed system is generated for fashion goods. The system is capable of generating persuasive slogans using specialized language from the fashion industry [12]. Tomaˇšič's proposed system can generate merchandising slogans with metaphors [18].

However, in the previous studies, the input was only text. The information provided by product images was not used. Our study uses product description as text, and at the same time uses product images as another input, using associative expressions that match the appearance of the product. The study on associative representation of images by Özbal et al associated proverbs from images [14]. Stampoulidis's study proposed a model that can distinguish metaphorical street art images from other rhetorical figures. the model seemly has the ability to identify metaphor from images [16].

3. Visual Metaphor

In our study, the most important issue is the matching of visual metaphor. In this section, we discuss the definition of visual metaphor, and the three visual elements for establishing visual metaphor. In addition, we describe what kinds of things make good visual metaphors.

We conclude in this section that the metaphors used in this study are defined as things that meet the following three conditions.

- Something that has a common impression in the public mind.
- Something that is similar in appearance to fashion goods.
- Something that reflects the charm of fashion goods.

3 1. Definition of Visual Metaphor

Visual metaphors need to have a general impression because they need to have a specific "appearance" to compare with fashion goods. This general impression is a single one. If a thing is perceived in many different ways by the general public, we will not be able to select any one of them for comparison. For example, "magic" looks different in everyone's mind. Some people think of red flames, while others think of brilliant beams of light. When we use "magic" as a metaphor for a red clothes, those who associate it with flames will recognize it, but those who associate it with colorful beams of light will not. Therefore, using a metaphor for something that has a diversity of impressions is not universally understood.

visual metaphors are based on the similarity of appearance to fashion goods. Therefore, metaphor words should be substances, or abstracts with common impression. As for latter, there is an example to explain. People may think about shiny blue sky and swimming pool when they imagine summer. When comparing fashion goods with the metaphor word "summer", it is same as comparing fashion goods with "sky" or "swimming pool".

3 2. Three Visual Elements of Visual Metaphor

In this study, we used **color**, **shape**, and **texture** as the three independent visual elements to measure similarity of fashion goods and metaphors. Some examples of visual metaphors founded on three visual elements are given in Figure 2. We predict the similarity of fashion goods to

Figure 2 Examples of visual metaphors derived from three visual elements

each metaphor for each visual element separately, listing the metaphors that are similar in terms of color, shape, and texture.

3 3. Conditions of Good Visual Metaphors

In Section 3 1 , we gave two properties that a visual metaphor needs to have. However, this does not mean that anything can be a good visual metaphor. For example, "printer" has a general impression and is similar in appearance to a fashion goods. When we use the word "printer" as a metaphor for this fashion goods, the advertising campaign is predictably bad. In order to select visual metaphors that achieve good advertising results, it is necessary to identify the properties of a good visual metaphor in addition to those defined in Section 3 1.

When we use sakura as a metaphor for a pink dress, we think the dress is cute. When we use the sea to compare a blue dress, we will think the dress is beautiful. These two examples are successful in advertising campaigns because we want our fashion goods to be cute and beautiful. In our impression, sakura are lovely and the sea is beautiful. When using sakura as a metaphor for fashion goods, the cuteness of sakura is given to fashion goods. When sea is used as a metaphor for fashion goods, the beauty of sea is given to fashion goods. This makes it possible to discover the commonality between fashion goods and visual metaphors.

In summary, properties of a fashion goods that excites consumers to buy it, are that a good visual metaphor needs to have. For example, beautiful, cute, handsome, etc.

4. Dataset

The datasets we use in this study are **fashion goods dataset**, **visual modifiers dataset**, **metaphor dataset**, and **one-word rhetoric dataset**. fashion goods dataset is extracted from Rakuten Ichiba Dataset [注1]. Metaphor dataset and visual modifiers dataset are collected from Internet. In **fashion goods dataset**, we mainly discuss about the method for filtering the categories of fashion goods and downloading images of fashion goods. In **metaphor dataset**, we would mainly discuss collection of metaphor images. In **visual modifiers dataset**, we would discuss the definition of visual modifiers and the collection of images. In **one-word rhetoric dataset**, we would discuss the collection of one-word rhetoric data.

4 1. Fashion Goods Dataset

We extracted data from Rakuten Ichiba dataset. The data in Rakuten Ichiba dataset are include:

- Product Name
- Store Code
- Product Code
- Product URL
- Price
- Category ID
- Image URL
- Sales Description
- Product Description

When performing catchcopy generation, we use only two of these data as follows.

- Image
- Product Description

First, we need to filter the goods that fall into fashion category from the entire data set. We use category ID to filter.

We extracted data whose parent category is in **women's fashion**, **men's fashion**, **bags, accessories and branded goods**, **shoes**. We exclude data except "bags" and "hats" from "bags, accessories and branded goods" and "shoe care and accessories" from "shoes" because they are functional goods. Then, we exclude data that in set, others and lucky bags. Finally, There are 218,585 pieces of data in 297 categories.

As **image URL** in dataset is often invalid, we filter from fashion goods dataset for those image URL is available. We get 103,737 pieces of data and define it as **fashion goods dataset**.

4 2. Visual Modifiers Dataset

Visual modifiers are words used to describe colors, shapes, and textures, as defined in this study alone. The visual modifier dataset used in this study contains 26,058 fashion images labeled with 33 visual modifiers. The visual modifiers defined in this study are shown in Table 1. Visual modifier images are collected by querying "[modifier] fashion" as a query and downloading the images through the Bing Search API. For each visual modifier, we collected 400 to 1000 fashion images. There are 15 labels for visual modifier data about color, and 11,562 images in total. The data about shape has 10 la-

(注1):https://rit.rakuten.com/data_release/

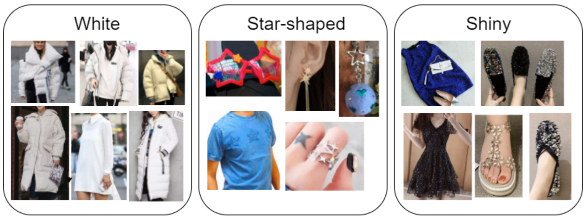

Figure 3 Examples for visual modifier images

Table 1 Visual modifiers of three visual elements for visual metaphor

visual element	visual modifier
color	black, white, red, blue, brown, yellow, orange, silver, pink, purple, gray, gold, cream, aqua, navy blue,
shape	round, sharp, triangular, square, star-shaped, cloud-shaped, heart-shaped, wide, thin, prickly,
texture	shiny, bling, rough, fluffy, slippery, smooth, tough, silky

bels and 8,044 images. The data about texture has 8 labels and 6,452 images. Figure 3 shows some examples of visual modifier images.

In this study, we focused on finding metaphors that are similar to the appearance of fashion goods. We would build an image recognition model to determine whether fashion goods are similar to metaphors to get metaphors that are similar to fashion goods. Visual modifier dataset is built with the purpose of providing training data for this model.

4 3. Metaphor Dataset

Metaphor dataset contains several metaphor word labels and 35 images under each label. We have artificially defined 100 metaphor words according to definitions in Section 3 1. Table 2 shows the 100 metaphor words.

Finally, we collected 35 images for each metaphor word separately using Bing Search API.

4 4. One-word Rhetoric Dataset

One-word rhetoric dataset contains a large number of one-word rhetorics for 100 metaphors. There are more than 150 one-word rhetorics under each metaphor. For example, rhetorics of the metaphor word “cookie” include “cookie of memories”, “cookie of hearts”, “cookie of Germany”, etc. These rhetorics were collected on the web through Bing Search API.

As for method of collecting one-word rhetorics, First, we searched for the specified query using Bing Search API to get snippets of articles on metaphor words. The query we use consists of metaphor word and **cue characters**. The purpose of using cue characters is to accurately retrieve snippets containing verbs, nouns, adverbs, and adjectives from large number of web search results. Next, we extracted verbs, nouns, adverbs and adjectives near the metaphor word from snippets as rhetorics of metaphor word.

Table 3 shows some examples of one-word rhetorics.

Table 2 100 metaphor words used in this study

metaphor words
sakura, plum, roses, sunflowers, dandelion, tulips, hydrangea, pine tree, bamboo, flower petals, strawberry, cherry blossom, mandarin orange, apple, lemon, chestnut, ice cream, cake, caramel, candy, cookies, marshmallows, chocolate, parfait, cocoa, coffee, soda, milk, wine, ribbons, lipstick, jewels, diamonds, crystals, rubies, emeralds, shells, pearls, sapphires, spring, summer, autumn, winter, universe, sun, star, shooting star, moon, full moon, crescent moon, aurora borealis, seashore, sandy beach, sea, deep sea, waterfall, garden, grassland, desert, waves, morning sun, sunlight through trees, sunny, sunset, night, fog, snow, powdered snow, rainbow, darkness, lightning, angel, fairy, spirits, dragons, bubbles, pyramid, fireworks, robots, rockets, wind chimes, blood, puddle, flame, ice cream, swallows, tiger, cat, dog, chick, elephant, giraffe, fireflies, butterflies, swans, rabbits, sheep, seagulls, lions, whale

Table 3 Examples for one-word rhetorics of metaphor words

metaphor word	rhetoric candidates
strawberry	loving strawberry, proud strawberry, exquisite strawberry, best strawberry, strawberry of night, strawberry of town, wild strawberry, delicate strawberry, felt strawberry
cherry	cherry in early summer, delicious cherry, sweet cherry, glittering cherry, special cherry, planted cherry, harvested cherry
cocoa	luxurious cocoa, warm cocoa, winter cocoa, various types of cocoa, soft cocoa, conspicuous cocoa, cocoa with black beans, cocoa with heart
cake	montblanc cake, celebration cake, recommended cake, christmas eve cake, today’s cake, pink cake, wonderful cake

5. Method

As showing in Figure 4, our approach is including **metaphor search** and **rhetoric generation**. First, we use the image of fashion goods to search the similar metaphors. Second, we use obtained metaphor word and product description to generate rhetoric part to complete catchcopy. In this section, we will discuss about the process of metaphor search and rhetoric generation.

5 1. Metaphor Search

We match metaphor to fashion goods by the approach

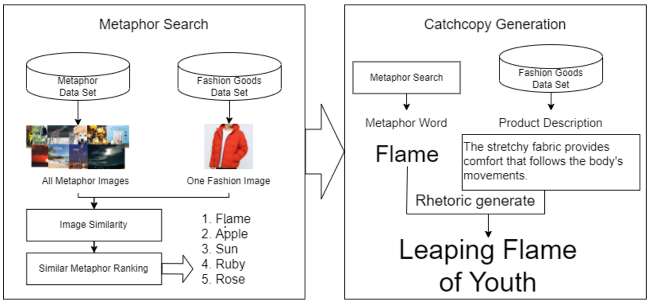

Figure 4 Two steps of fashion goods catchcopy generation

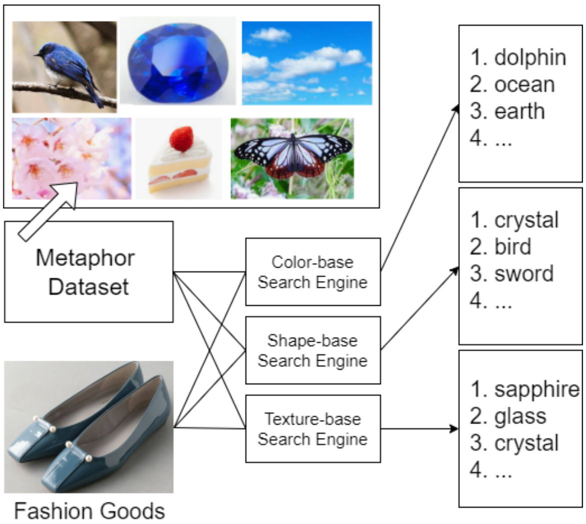

Figure 5 Visual metaphor matching process based on three metaphor search engines

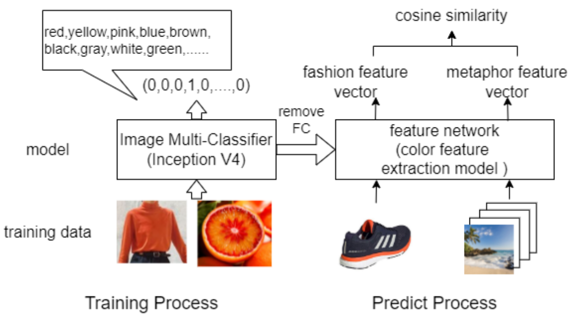

Figure 6 Training process and prediction process of color feature extraction model

showed in Figure 5. We predict the similarity between image of fashion goods and each metaphor in metaphor dataset. The similarity prediction is based on three visual elements: color, shape, and texture. The ranking of similar metaphors is obtained based on each of these three visual elements.

In this section we mainly explain the process of matching similar metaphors for fashion goods. This includes the approach of building three metaphor search engine.

5 1. 1. Metaphor Search Engines

In order to implement three metaphor search engines based on color, shape, and texture, we have to build three feature extraction models to obtain the feature vectors of images in terms of color, shape, and texture. we use Inception V4 [17] to train three models for obtaining image's feature vector on color, shape and texture. In this study, we define these three models as color feature extraction model, shape feature extraction model, and texture feature extraction model. Figure 6 shows the building of feature extraction models. We first constructed three image classifiers. The labels of classifiers are defined as same as the words in Table 1. There are 15 classes for image multi-classifier of color, There are 10 classes and 8 classes for shape and texture.

Our aim is that each classifier is able to distinguish different images by each visual element. Color-base classifier can classify images based on their primary color. The shape-base, on the other hand, classifies images based on the shape of objects in the image. texture-base classifies images based on the textural patterns that appear in the images. For example, for color-base classifier, pink dresses and peaches can be classified in same class. Black shoes and a crow would be identified as same class.

However, our ultimate goal is not to classify the images, but to get the feature vectors of the images. We obtain images' similarity by calculating cosine similarity of the feature vectors of the images. In Figure 6 we illustrate the process with color feature extraction model as an example. First, we train image multi-classification model of color. We discard its fully-connected layer to make it a feature network capable of pumping out feature vectors. We perform feature vector extraction on 100 metaphors prepared in advance. The extraction process of the metaphor feature vector is shown in Figure 7. First, we extracted the feature vectors of 35 images of the same metaphor, and then averaged them. We define the final vector as the feature vector of the metaphor. Then we take fashion image as input and get the feature vector of it. Finally, we calculate the cosine similarities of fashion feature vector and all of metaphor feature vectors. The calculated cosine similarities is sorted in descending order, and the ranking of visual metaphors of fashion goods is obtained.

5 1. 2. Training Data

In this section, we illustrate the training data for image multi-classifiers. The training data for the models is defined as fashion images with various features. We use visual modifiers dataset mentioned in Section 4 2 as the training data.

For each training process of three classifiers, We partition the training data in the ratio of 8:1:1 to obtain the training set, validation set and test set. Training set for classifier of color has 9,249 images. Validation set has 1,156 images and test sets has 1,157 images. Classifier of shape's training data has 6,435 images in training set, 804 images in validation set and 805 images in test set. Classifier of texture's training set

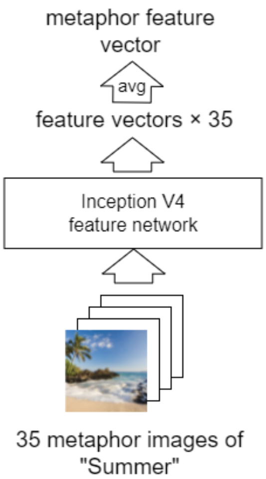

Figure 7　Extraction of metaphor feature vector

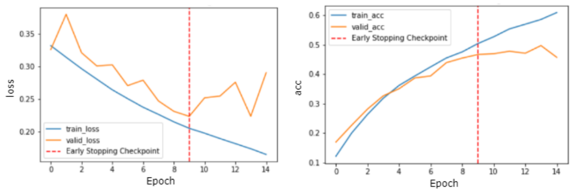

Figure 8　Learning curve of color-base classification model

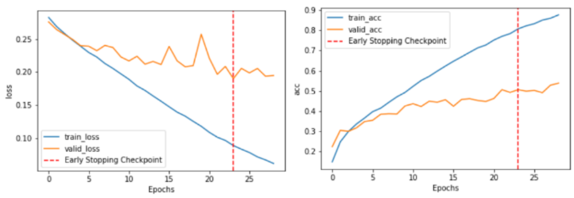

Figure 9　Learning curve of shape-base classification model

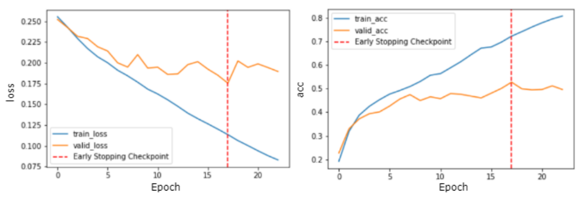

Figure 10　Learning curve of texture-base classification model

has 5,161 images. Validation set has 645 images and test set has 646 images.

5 1. 3.　Result

We trained three classifiers with the same following parameters.

- Batch Size：8
- Optimizer：Adam [11]
- Criterion：Cross Entropy Loss
- Learning Rate：2e-6

Figure 8,Figure 9 and Figure 10shows the training process of color classifier, shape classifier, texture classifier.

All three models showed that the validation set loss stopped decreasing before the training set loss decreased to a very small value. The learning processes of shape classifier and texture classifier also appear to be over-fitting. Therefore, as shown in Table 4, all three models have low accuracy on test set. That is, the three models trained cannot achieve accurate classification of fashion images according to our prescribed visual modifiers.

Table 4　Accuracy of visual feature classification models

model	accuracy
color-base	0.45
shape-base	0.54
texture-base	0.57

Figure 11　Example for output of color-based metaphor search

However, this result does not affect the performance of metaphor search engine. The reason for this is that our ultimate goal is not to classify images accurately, but to obtain their feature vectors based on color, shape and texture. The reason for low accuracy of constructed models is that the training data have similar images even though they belong to different classes. For example, the image with label “navy blue” is similar to the image with label “blue”. This makes it difficult for the model to perform accurate classification.

An example of metaphors obtained using metaphor search engine is shown in Figure 11. This example is about output of color-base metaphor search engine. The input image is a pair of brown shoes. The top 10 matching metaphors are chocolate, pyramid, dragon, swallows, desert and so on.

5 2.　Rhetoric Generation

In this section, we will generate rhetoric based on metaphor word to complete the catchcopy. The flowchart of rhetoric generation is shown in Figure 12. The rhetoric part of catchcopy has two words. To obtain the first word, we used the dataset mentioned in Section 4 4 to rank all rhetorics of metaphor word based on two scores. We select the first ranked rhetoric and generate the second word of rhetoric part by MASK estimation.

5 2. 1.　Matching of First Rhetoric Word

In this section, we calculate two scores to rank all one-word rhetorics of the metaphor word, and then select the one with the best score as the first word of rhetoric part. The first

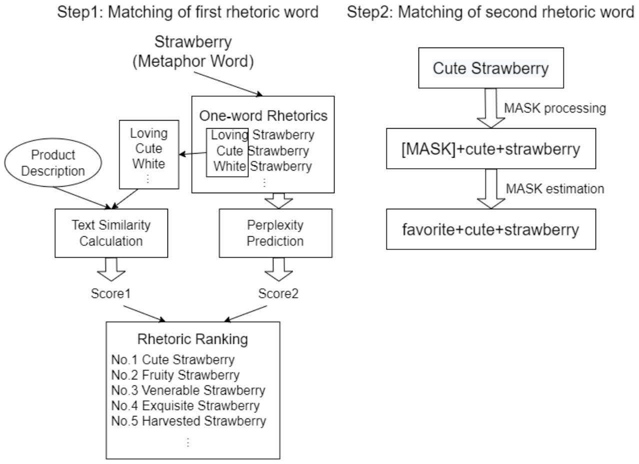

Figure 12 Process of rhetoric generation

score is textual similarity to product description of fashion goods. The second score is the perplexity(PPL) taken in reverse. Textual similarity is calculated by vectorizing product description and one-word rhetoric removing metaphor word using pre-trained BERT [9], and then calculating their cosine similarity. Perplexity defined in this study is the value of the metric text naturalness obtained by GPT-2. Its value domain is $[0, \infty]$. Since the value of perplexity is a negative logarithm, the smaller the value, the more natural the text is. We perform min-max normalization on the two scores. Outliers of perplexity affect the normalization to a large extent, so that the distribution after normalization is too dense to distinguish each value well. We solve this problem by converting all the values of perplexity greater than 2500 to 2500. We bring the normalized two scores into the following equation to calculate the final match for the first rhetoric word. We define the final match of rhetoric word as M, Textual similarity between rhetoric word and product description as S, and the perplexity of rhetoric as P.

$$M = 0.5 * S + 0.5 * P$$

Finally, we sort the final matches in descending order and select the rhetoric word with the best final match as the first word of rhetoric part.

5 2. 2. Prediction of Second Rhetoric Word

we perform MASK estimation in order to get the second word of the rhetoric of catchcopy. We add MASK and a cue character in front of the first rhetoric word and make prediction for it. Cue character can be Japanese words ta, no, na. Figure 13 shows the complete output flow for generating catchcopies with visual metaphor based on fashion goods.

6. Case Study

In this section, we will examine several specific output examples of catchcopy generation. Figure 14 shows some examples of catchcopies generated by our approach on color-base, shape-base and texture-base.

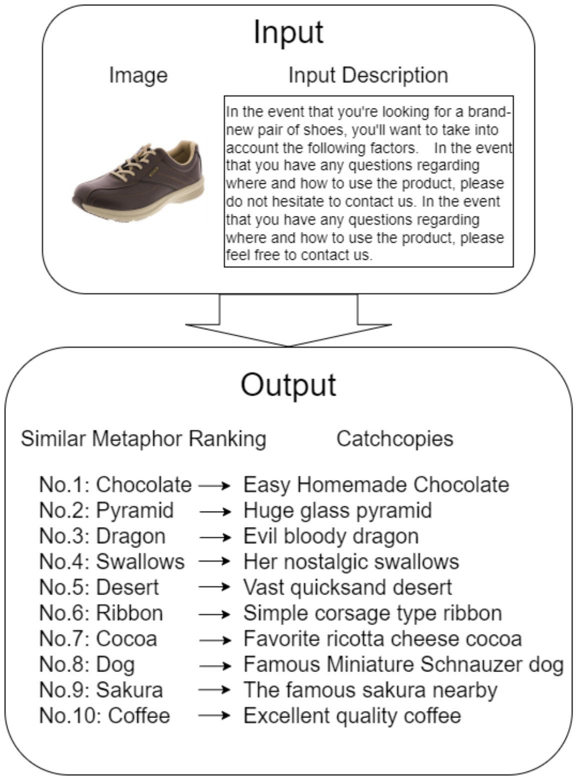

Figure 13 Catchcopy generation for fashion goods on color-base

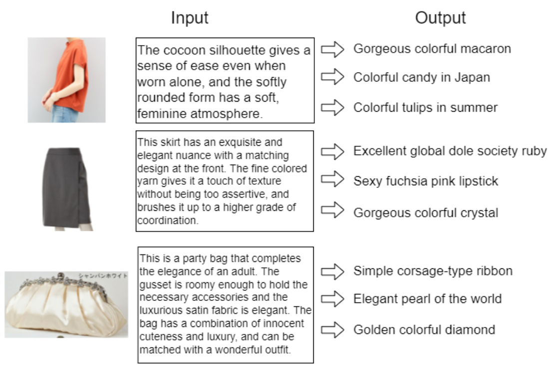

Figure 14 Catchcopy generation for fashion goods on color-base, shape-base, texture-base

Three catchcopies were created for a soft and feminine orange top, each containing the visual metaphors “macaroon”, “candy” and “tulip”. Each of the three was very appropriate to fashion goods. However, the word “colorful” is used repeatedly in rhetoric. Moreover, fashion goods is monochromatic and does not fit the term “colorful”.

For a delicate and elegant black skirt, the three catchcopies were created using a shape-based approach, each containing the visual metaphors “ruby”, “lipstick”, and “crystal”. The three metaphors do fit the shape of fashion goods. However, the image of ruby and lipstick is red, while the color of the skirt is black. It is not good for using ruby or lipstick as a metaphor for this case, even though they have similar shapes.

For a high class handbag, the three catchcopies obtained using texture-based approach contain visual metaphors "ribbon", "pearl", and "diamond". Texture-based metaphor search engine accurately captures the silky and shiny texture of the handbag.

7. Conclusion

This paper proposes a system to automatically generate catchcopy from fashion goods images and description. As for research process, we first defined the properties that a visual metaphor needs to have and what a good visual metaphor is, and artificially prepared 100 metaphor words. The generation of a catchcopy for a fashion goods consists of two steps. First, in metaphor search session, we use the similarity between fashion goods image and metaphor image to get the ranking of similar metaphors. To obtain similar metaphorical rankings, we trained three image classification models separately to extract color, shape, and texture feature vectors of images. By calculating the cosine similarity between fashion goods feature vector and metaphor feature vector one by one, we obtain the 10 most similar metaphors to the fashion goods. Then, in rhetoric generation session, we generate rhetoric for each metaphor word in the ranking to complete the final catchcopy. First, we collected a large number of single rhetoric word for each metaphor words from Internet as candidates. By calculating the textual similarity between product description and rhetoric candidates, we get one rhetoric word that best matches the product information. Finally, another word of rhetoric part is predicted by MASK.

We consider making an evaluation of the results of this study in the future. At present, we initially intend to conduct the evaluation in the form of a questionnaire distributed to users. The questionnaire is designed to rate three aspects of the generated catchcopies. The three aspects are fluency, relatedness, and novelty. The implementation of the questionnaire will be considered in the future.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers JP21H03775, JP21H03774, JP21H03554. In this paper, we used "Rakuten Dataset" (`https://rit.rakuten.com/data_release/`) provided by Rakuten Group, Inc. via IDR Dataset Service of National Institute of Informatics.

References

[1] Khalid Alnajjar and Hannu Toivonen. Computational generation of slogans. *Natural Language Engineering*, 27:575–607, 2021.

[2] Swee Hoon Ang and Elison Ai Ching Lim. The influence of metaphors and product type on brand personality perceptions and attitudes. *Journal of Advertising*, 35:39–53, 2006.

[3] Sandy Bulmer and Margo Buchanan-Oliver. Visual rhetoric and global advertising imagery. *Journal of Marketing Communications*, 12:49–61, 2006.

[4] Shuo Cao, Huili Wang, and Xiaoxia Zou. The effect of visual structure of pictorial metaphors on advertisement attitudes. *International Journal of Marketing Studies*, 10:60–72, 2018.

[5] Forceville Charles. Pictorial metaphor in advertisements. *Metaphor and Symbolic Activity*, 9:1–29, 1994.

[6] Charles Forceville. Visual and multimodal metaphor in advertising: Cultural perspectives. *Styles of Communication*.

[7] Charles Forceville. Pictorial metaphor in advertisements. *Metaphor and Symbolic Activity*, 9:1–29, 1994.

[8] Kango Iwama and Yoshinobu Kano. Japanese advertising slogan generator using case frame and word vector. In *Proceedings of the 11th International Conference on Natural Language Generation*, pages 197–198, 2018.

[9] Devlin Jacob, Chang Ming-Wei, Lee Kenton, and Toutanova Kristina. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, pages 4171–4186, 2019.

[10] Se-Hoon Jeong. Visual metaphor in advertising: Is the persuasive effect attributable to visual argumentation or metaphorical rhetoric? *Journal of Marketing Communications*, 14:59–73, 2008.

[11] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proceedings of the 2015 International Conference on Learning Representations*, 2015.

[12] Vitobha Munigala, Abhijit Mishra, Srikanth G. Tamilselvam, Shreya Khare, Riddhiman Dasgupta, and Anush Sankaran. Persuaide! an adaptive persuasive text generation system for fashion domain. In *Companion Proceedings of the The Web Conference 2018*, pages 335–342, 2018.

[13] Gözde Özbal, Daniele Pighin, and Carlo Strapparava. Brainsup: Brainstorming support for creative sentence generation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 1446–1455, 2013.

[14] Gözde Özbal, Daniele Pighin, and Carlo Strapparava. A proverb is worth a thousand words: learning to associate images with proverbs. In *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, pages 2515–2521, 2019.

[15] Barbara Phillips and Edward Mcquarrie. Beyond visual metaphor: A new typology of visual rhetoric in advertising. *Marketing Theory*, 4:113–136, 2004.

[16] Georgios Stampoulidis and Marianna Bolognesi. Bringing metaphors back to the streets: a corpus-based study for the identification and interpretation of rhetorical figures in street art. *Visual Communication*.

[17] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, page 4278–4284, 2017.

[18] Polona Tomašič, Martin Žnidaršič, and Gregor Papa. Implementation of a slogan generator. In *The Fifth International Conference on Computational Creativity*, pages 340–343, 2014.

[19] Haoxing Zhao and Xiaoyong Lin. A review of the effect of visual metaphor on advertising response. In *Proceedings of the 4th International Conference on Economy*, pages 29–34, 2019.