Journal of Data Intelligence, Vol. 4, No. 1&2 (2023) 134–148

AUTOMATIC GENERATION OF PRODUCT DESCRIPTIONS USING DEEP LEARNING METHODS

AKIYO NADAMOTO
Konan University, 8-9-1 Okamoto Higashinada-ku
Kobe, Japan
nadamoto@konan-u.ac.jp

KENJI FUKUMOTO
Konan University, 8-9-1 Okamoto Higashinada-ku
Kobe, Japan
m2124005@s.konan-u.ac.jp

RISA TAKEUCHI
Konan University, 8-9-1 Okamoto Higashinada-ku
Kobe, Japan
s1871056@s.konan-u.ac.jp

HIROYUKI TERADA
Ochanoko–net Inc., 7–1–1 Kumoidori Chuo–ku
Kobe, Japan
terada@ocnk.net

MASAFUMI BATO
Contact Inc., 2–17–1, 7–1–1 Kumoidori Chuo–ku
Kobe, Japan
bato@contact.co.jp

With flea markets becoming popular on the internet, people have begun to post their own products. In order to make their products more attractive, people(sellers) post a description of their products at the same time. However, it is difficult for ordinary people to write attractive product descriptions. In this study, we propose a method for automatically generating the product description based on a comparison of LSTM, GPT-2, and GPT-2 Rinna model. Specifically, we propose the data structure of the product included in the existing product description. Then we use deep learning models to generate attractive product descriptions using input sentences based on the data structure of the product. Furthermore, we conduct two kinds of experiments to measure the benefits of our proposed methods based on our proposed evaluation axis.

Keywords: Content generation, Deep learning, Product description, LSTM, GPT-2, GPT-2 Rinna-model

1. Introduction

With flea markets becoming popular on the internet such as Poshmark,[a] Mercari,[b] and so on, people post their own products. Generally, when people post their products on flea markets,

[a]https://poshmark.com/
[b]https://jp.mercari.com/

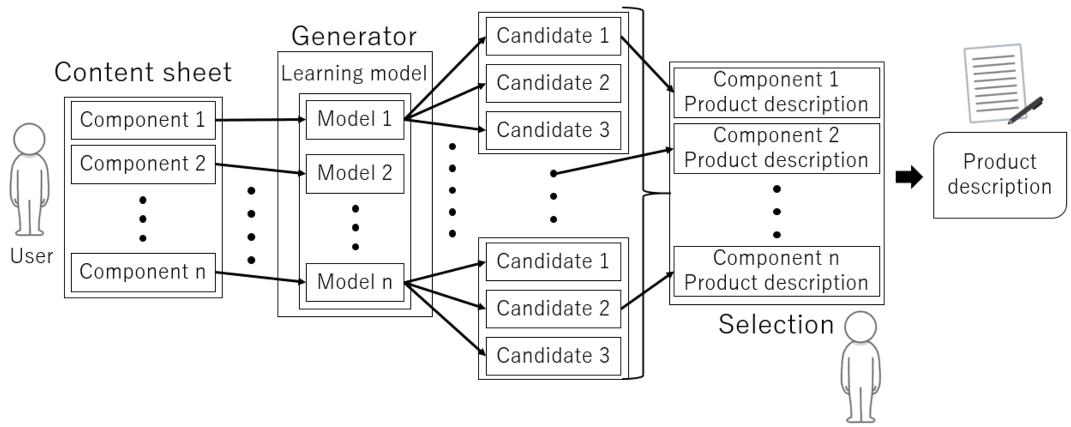

Fig. 1. Product description generation flow.

the sites require them many steps such as measuring products, creating images, creating product pages, and shipping. When a seller generates a product page, the text requires not only basic information such as the product name but also sentence(s) that explain the product details. In this study, the sentences that describe the product are called "the product description". A seller who is not a professional seller is called a "user". The product description describes features that cannot be explained by visual information such as product images, which engenders an increase in the purchasing motivation of E-commerce site consumers. As the number of flea market-style sites on E-commerce sites increases, general users have also become able to sell their own products. Along with this, it is necessary to describe the product description, but it is difficult for general users to create the product description from scratch. Therefore, for this study, we propose a method for automatically generating product descriptions. Specifically, we propose the method of automatic generation of the product description based on a comparison of Long Short-term Memory (LSTM)[1], the Generative Pre-trained Transformer 2 (GPT-2)[2], and GPT-2 Rinna model.[c]

When automatically generating the product description, using only the product name as input is unrealistic because each product has its own characteristics. The characteristics of each product consist of components that make up the product, such as "This chair is the color black, the material is leather, and the reclining can be done in five stages." The existing product description on the E-commerce site is written for each product component. Therefore, we generate and present the product description for each component of the product. The user selects the most suitable sentences from the generated product description for each structure and subsequently completes the product description for that product. Therefore, we create a data structure of the component of the product and propose "the content sheet," which is an input sheet based on the data structure. Then, we propose a method to generate a product description for each component on the content sheet. Figure 1 presents an image of the system flow.

In addition, E-commerce sites do not handle products of many types. Therefore, we do

[c]https://github.com/rinnakk/japanese-pretrained-models

not target all products, but instead generate product descriptions for furniture because of the appearance and size of clothes, which are important information obtained from product images. Therefore, the image is more important than the item description. For personal computers and home appliances, the specification table is often more important than the explanation because the usability and product features that cannot be conveyed from the product image described in the product description are important for when one makes a purchase. Therefore, as described herein, we propose an automatic generation method for product descriptions for furniture. We believe that our proposed method will reduce the labor necessary to create product descriptions.

As described in this paper, we contribute following the three points.

- Proposing a content sheet that consists of the data structure of products.
- Proposing the methods of automatic generation of product description based on the content sheet.
- Proposing the three indicators of broken grade, correctness, and diversity to measure the automatically generated product description, and conduct experiments.

This paper is organized as follows: Section 2 discusses related work; Section 3 proposes a content sheet; Section 4 proposes two types of automatic generation of product description methods; Section 5 presents subjective experiments, and Section 6 presents an objective experiment. Finally, Section 7 presents the conclusions of this paper.

2. RELATED WORK

Many studies have positively examined content generation using the Deep Learning model. Cao et al.[3] propose table-to-text generation using LSTM. Lee et al.[4] propose Recipe-GPT, which extends the GPT-2 model for automatic recipe generation. Jieh-Sheng [5] proposes the method of generating patents using GPT-2. Parakhar et al.[6] used the pre-trained Transformer Language Model to generate titles automatically from short texts. Zhou et al. [7] proposed a Multi Mode Description Generator (MMDG) to generate a point of interest (POI) description on the map. It consists of a multi-mode encoder and a Transformer-based-decoder to generate descriptions of POI. Zhang et al.[8] proposed a description generation model based on related and non-related documents for query comprehension. Their proposed model can understand the query well through detailed and accurate explanations. They use the deep learning method of text generation. It is similar to our study, but our research target generating product description differs from theirs.

Many studies have examined the generation of product descriptions. Tao et al.[9] propose a pattern-controlled neural model for generating product descriptions. Their proposed method uses the paradigm of a pointer–generator mechanism to resolve difficulties related to constraints such as repeatability and accuracy in the automatic generation of product descriptions. Wang et al.[10] propose a text generation model for product descriptions. The model is based on the user's click behavior related to previously generated content as context. Kedia et al.[11] propose an eBERT that generates voice titles in voice-based e-commerce. They used it to generate concise and grammatically correct titles. Novgorodov et al.[12] propose a method that generates product descriptions based on reviews Chen et al. [13] propose a method that generates product descriptions based on the transformer. Their method includes feature characteristics to the generated product description. The target of their methods is generating product descriptions, which is similar to our research. However, the point of our research is to generate a product description that includes the data structure of the product. This is a point that differs from other studies. Taira et al. [14] propose a method for analyzing TV commercials and generating product descriptions on a rule basis. They generate a content sheet by layering all the information about the product from the utterance content of the CM. They generate a product description from this content sheet. Our content sheet is based on their content sheet. However, whereas they use the structure of the product used in the commercial, we particularly examine the structure of the product itself to generate the content sheet. This is a different point from their proposed method to our method.

3. Content Sheet

Product descriptions on E-commerce sites are often written based on the structure of product attributes such as product type, material, and size. Therefore, the proposed product description automatic generation system generates a description based on the data structure of the attributes of this product. Then, we analyze the existing product description to determine the data structure of the product.

As shown in Table 1, each sentence of the product description is written for each attribute of the product information. Attributes are the type, material, size, etc. of the product. Especially in the case of furniture, the attributes of product information have a hierarchical structure. Therefore, in this research, we create a data structure of attribute information for each product and create a content sheet based on that data structure. In this paper, the data structure of attribute information is called the attribute information data structure. The attribute information data structure is created by manually analyzing the page structure and product description of the existing E-commerce site.

In this paper, we targeted the furniture in the furniture and created the attribute information data structure of the sofa. We analyzed 5 E-commerce sites which are Specifically, BELLE MAISON [d], FURNITURE DOME [e], IKEA [f], LOWYA [g], nissen [h]. We analyzed the page structure of 5 E-commerce sites and the product description of 1,760 and we created the attribute information data structure of the furniture. As a result of the analysis, we found that the data structure of furniture consists of "a common data structure of furniture" and "a specific data structure of the type of furniture." At this time, prices, etc. not described in the item description were deleted from the attribute information data structure.

The common data structure of furniture consists of common components such as color, design type, target user, and so on. The specific data structure of the type of furniture differs from furniture type. Figure 2 shows the common data structure of furniture and Figure 3 shows the specific data structure of the sofa. These data structures consist of three layers. The second layer below the root node is furniture attribute data. We designate the second layer "component." The third layer is an instance of each component. When we create the

[d] https://www.bellemaison.jp
[e] https://www.furnituredome.jp
[f] https://www.ikea.com Five EC sites: / jp / ja /
[g] https://www.low-ya.com
[h] https://www.nissen.co.jp/s/interior/

Table 1. Examples of sofa descriptions

Component	Product description
Design	The design is simple and fits into various styles.
	The two-seat sofa features a retro design.
Type	The sofa is a couch sofa that allows you to relax.
	The sofa is a steplessly adjustable reclining sofa; you can change the backrest angle freely.
Cushion material	The seat cushion is pure urethane and a highly durable pocket coil.
	The seating surface, which is exhausting if it is too soft, is urethane with moderate elasticity.
Surface material	It uses synthetic leather with excellent durability on the surface that does not touch the body when sitting.
	The sofa uses a soft woven fabric.

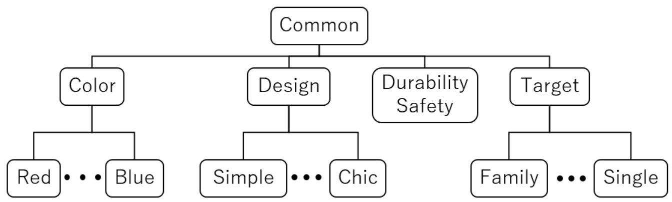

Fig. 2. Common data structure of furniture.

data structures, prices are not described in the product description. Then we deleted the price from the component in the data structure.

When generating the product description based on the data structure, not all components are important for a product description. Therefore, we propose a system in which the user inputs only the data of the required component. The system generates a product description for each input component. We also propose a content sheet, which is a sheet input by a user. Figure 4 presents an example of our proposed content sheet. From Figure 4, the content sheet is based on the data structure. The user inputs only the value of the component that the user wants to put in the item description. As a result, the system generates the product description of the input component.

4. Generating Product Description

We generate a product description for each component described in the content sheet. Figure 1 shows the image of generating product description. We propose three methods for generating product descriptions, LSTM, GPT-2, and GPT-2 Rinna model.

4.1. *Training data*

As training data for deep learning, we extract product descriptions for sofas from the EC sites of 36 companies that sell furniture. We obtained 4,744 documents from these sites. The document consists of 34,035 sentences. The furniture description describes basic product

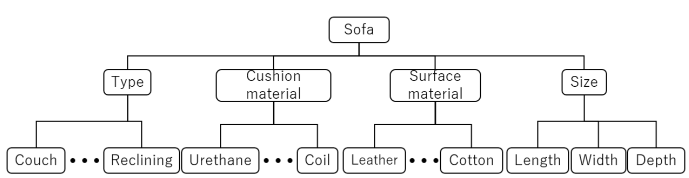

Fig. 3. Specific data structure of a sofa.

Content sheet		
Furniture type	Sofa	
Common data structure		
Color		
Design		
Durability		
Target		
Specific data structure for sofa		
Type		
Cushion material		
Surface material		
Size	Length	
	Width	
	Depth	
Free comments		

Fig. 4. Sofa content sheet.

information and product strengths. Furthermore, some sentences enhance the reliability of the product by explaining in detail the company or person who made the product. However, in the product description, a specific company name or person name differs depending on the product. The user should then input them. Furthermore, the numbers included in the product description differ depending on the product. The user should input them. Therefore, for this study, we mask proper nouns and numbers in advance as data pre-processing. The masking rules are presented below.

- Proper noun
  We perform masking to delete the company name and personal name included in the product description. We change all words for which the part of speech is a "proper noun" and for which the part of speech subclassification is not "common" to <pnoun>. For example, "The sofa is produced by Takano Factory Co., Ltd." is an original sentence. Then we change the original sentence to "The sofa is produced by <pnoun>."

- Number
  We mask the notation of numbers such as the size of the product in the product description. Specifically, we mask numbers with <num>using regular expressions. As an example, "The sofa is for three people, uses five colors, and has an eight-stage reclining seat" is an original sentence. Then we change the original sentence to "The sofa is for <num>people, uses <num>colors, and has an <num>-stage reclining seat."

We used MeCab [i]for morphological analyses used to determine proper nouns, and mecab-ipadic-Neologd [j]for the dictionary.

4.2. *LSTM Method*

First, we propose the use of LSTM to generate product descriptions. Therefore, LSTM can assess expressions that can not be evaluated using other methods based on term sequences.

[i] http://taku910.github.io/mecab/
[j] https://github.com/neologd/mecab-ipadic-neologd: mecab-ipadic-NEologd

We use Python's machine learning library Chainer [k] to implement LSTM.

We use vectorization of all the words appearing in the tweet using fastText[l]. At this time, we use NWJC2Vec[15] as a Japanese fastText model. The feature vector words are all parts of speech in tweets. The reason for using all parts of speech is that we infer that sequences must be considered in the product description. We also use 0 vectors for unknown words that do not exist in the model we used. When we convert the result back to a word, we convert it to UNK. We ascertain the various LSTM parameters using a grid search. The parameters are 1 for the hidden layer, 168 for the unit number, 1024 for the batch size, 100 for the epoch number, 0.001 for the learning rate, and 8 for the maximum word length. The optimizer uses Adam[16]. Particularly, our LSTM model consists of one layer. The sequence length is the number of words in a tweet. The 300-dimensional feature vector is the input for the unit of the LSTM. The output of the LSTM unit becomes the input of the next LSTM unit. The last output of the LSTM unit becomes the input of the fully connected layer. The fully connected layer outputs two values. We use softmax to divide behavioral facilitation tweets or not.

4.3. *GPT-2*

GPT-2, a large-scale learning language model based on Transformer, consists of pre-learning and fine-tuning. Using it, learning is possible with small-scale data. Using this large-scale language model, we can expect more natural sentence generation than LSTM. For this study, we generate a new product description using a large-scale language model based on the original product description that has been preprocessed. The GPT-2 model used for this study is a Japanese model in which about 7 million web pages are learned in a 24-layer network using about 320 million parameters.[m] This study generates the product description using the training data that we described previously as fine-tuning.

We use Corpus 2020 [n] used for GPT-2 pre-learning. Corpus 2020 consists of 3.2GB of blog articles, 1.8GB of questions / summary / scoring site articles, 3.4GB of web dictionary site articles including Wikipedia, 2.1GB of news release articles, 0.3GB of news site articles, and the web. The articles on the novel site are 13.1GB. The number of articles is 6,755,346. The number of tokens is 3.86 billion.

We use 40, which is reported to be the best result, as the value of the number of candidates for predicting the next word from the top few used for hyperparameters. We use Japanese Byte-Pair Encoding (BPE) Encoder [o] as the Fine tuning model of GPT-2. Therefore, GPT-2 can learn word separation and dependence by division into finer characters than the word-separation used in conventional natural language processing.

4.4. *GPT-2 Rinna model*

The GPT-2 Rinna model is a model published by rinna corporation.[p] In this paper, we use the GPT-2 Rinna model to generate product descriptions based on the pre-processed product

---

[k] Chainer `https://chainer.org/`
[l] fastText `https://fasttext.cc/`
[m] `https://github.com/tanreinama/gpt2-japanese/blob/master/report/models.md`: GPT-2 Japanese model
[n] `https://github.com/tanreinama/gpt2-japanese/blob/master/report/corpus.md`: Corpus 2020
[o] `https://github.com/tanreinama/Japanese-BPEEncoder`: JapaneseBPEEncoder
[p] https://corp.rinna.co.jp/

Table 2. Example of input sentences

Attribute	Input sentence
Cushioning	The material of the seat is <nakazai>,
	We use the <nakazai>as a seat,
Design	the design is <design>
	The design of the sofa is <design>

descriptions. We used CC-100[q]as the training data for pre-training the GPT-2 Rinna model. We also used Huggingface Transformers[r]as fine-tuning.

For input, we create a sentence for each attribute that contains the keywords entered in the content sheet in advance. We input the sentences to the system for each attribute. The reason is that the language model is a model that predicts the next word sequentially from the input words. Then we consider that by using sentences containing keywords as input, we can generate sentences that correctly describe the product. In addition, we generate templates as the input sentences. We generate the templates manually from existing product descriptions on the Internet. The template consists of 7 items which are a type of product attribute, cushioning material, surface material, size, target audience, color, and design(show in Figure 4), with 3 sentences each, for a total of 21 sentences. The value of the content sheet is indicated by tags which are <nakazai>and <design>tags in the input text. Table 2 shows a part of the template for input sentences. When the system presents the product descriptions to the user, the system connects the input sentence and output sentence as one sentence.

5. Experiment1: Subjective evaluation of the generated product description

We conducted experiments of two types: subjective evaluation and objective evaluation. This section presents a discussion of the results of subjective evaluation. We subjectively analyzed the contents of the generated product description visually. Then we identified the difficulties inherent in our proposed methods. The evaluation data are 1,000 sentences, which are 500 randomly extracted sentences of each product description generated using the two proposed methods. Table 3 presents an example of the evaluation data.

5.1. *Analysis and Discussion of the LSTM results*

In product description generation using LSTM, the sentence was generated correctly. A sentence appealing to the product was generated. We use LSTM based on word-by-word. In product description generation using LSTM, the sentence was generated correctly. Sentences making an appeal about the product were generated. The reason is that LSTM is based on word-by-word. However, the system generated a sentence in which the same word appears repeatedly, as in LSTM-(3) in Table 3. The reason is that in our model, word prediction is eight words, which is small. Furthermore, we did not consider the appearance of the same word. We were unable to consider a penalty for the appearance of the same word. In addition, a sentence such as LSTM-(2) in Table 3 is generated in which the product description is followed by a description of sitting comfort for the subject of "depth of the seat". In this way, a sentence that makes you feel uncomfortable is generated as a product description.

[q]http://data.statmt.org/cc-100/
[r]https://github.com/huggingface/transformers

Table 3. Evaluation Data Examples

Method	Product Description
LSTM	(1) You can adjust the backrest angle to relax in various postures.
	(2) The depth of the seat is firm, so you can sit comfortably. In addition, the back cushion has plenty of feathers.
	(3) The design of the grid on the back is included in the design of the grid on the back. Therefore, there is no feeling of oppression even if it is placed in the center of the room. It does not feel oppressive even if it is placed naturally in the room.
GPT-2	(1) The sofa is processed like a deep fabric to give it a wooden texture and a natural atmosphere. It has a warm finish with a modern impression.
	(2) The seat surface with the number stage reclining function can recline in the number stage. Therefore, you can enjoy a relaxing and relaxing time even if you sit comfortably.
	(3) The arm is designed so that you can sit comfortably in the storage space under the seat surface on the seat surface that is not too high and not too low. You can relax forever.
GPT-2 Rinna model	(1) The color of the sofa is white, so you can coordinate it with a natural-taste low table or a natural table to create a stylish cafe style. Since it is a floor sofa, you can freely change the layouts. The modern design of the sofa makes it suitable for a wide variety of interiors, such as Scandinavian style, <pnoun> and <pnoun>. The seat of the sofa is made of cotton, so it can be used comfortably all year round. This sofa is recommended for families with children, so the whole family can get along and make memories together. The material of the sofa surface is fabric, so the cover can be removed.
	(2)The color of the sofa is white, which adds a light and airy atmosphere to the room. It is a reclining sofa, so you can relax in comfort. The modern design of the sofa is simple, so it goes well with furniture with various atmospheres, and it can be placed in a room in a variety of ways. The material used for the seat of the sofa is cotton, so it can be used throughout the year. This sofa is recommended for families, so please consider it. The surface of the sofa is made of fabric, so it is easy to clean.
	(3)The couches are available in brown and black, so you can choose the one that best suits your room. The type of sofa is a couch, so you can lay out the corners as well as the ottoman and both sides together. The design of the sofa is modern, but after all, a low sofa is the best way to enjoy the resort feeling while at home! I thought this <num>seat would be a great way to redecorate the room. The seat of the sofa is made of feathers. It feels warm and comfortable all year round. As the sofa is designed for men, the seat is wide and easy to use. The surface of the sofa is made of genuine leather, which has a matte and moist feel that only genuine leather can provide. The matte and moist feel of genuine leather further enhances the comfort of use.

We also specifically examined one word: "reclining". Reclining is the type in the specific data structure. Results show that among the 21 generated sentences including "reclining", 20 generated sentences similar to "backrest is <num>-step reclining" were found. From this result, similar phrases were used and were generated in word-based sentence generation using LSTM. Results illustrate the same case for other phrases. The results showed that using LSTM does not produce various product explanations, the reason is that similar phrases are used repeatedly. Therefore, studying a method for generating various phrases is necessary.

5.2. *Analysis and Discussion of GPT-2 results*

In sentence generation by GPT-2, a sentence "and has a warm finish with a modern impression." in GPT2-(1) in Table 3, and "so you can enjoy a relaxing and relaxing time even if you sit comfortably." in GPT2-(2) in Table 3, one can generate the appealing product description using GPT-2. However, for example in the GPT2-(3) in Table 3 "under the seat surface, on the seat surface that is not too high and not too low and you can relax forever.", the grammar of the sentence is correct, but the content is strange.

We also emphasized one word, “reclining” similarly to LSTM. From the generation results, there were 36 generated sentences that included the word “reclining”. There were a few generated sentences using similar phrases. Product explanations using various phrases were generated. This finding differs from those of LSTM. From this finding, we inferred that GPT-2 can generate more varied sentences compared to sentence generation using LSTM.

5.3. *Analysis and Discussion of GPT-2 Rinna model results*

The GPT-2 Rinna model has a much larger amount of sentences than the other models. It also generates a variety of attractive sentences. For example, when we focus on reclining as in LST and GPT-2, as shown in GPT-2 Rinna model(2), the sentences are not only describing reclining but also generate sentences that give us an image of using it, such as “You can relax and lie down”. Furthermore, for each of the various attributes of the product, sentences that evoke a feeling of use and an image of the product when placed in a room are generated. On the other hand, the sentences are longer than those of LSTM and GPT-2, so the same expressions are used, and some sentences are broken.

6. Experiment2: Objective evaluation of the generated product description

We consider that it is important for a product description to motivate people who are looking at the product (hereinafter referred to as ”browsing users”) to buy it. However, it is not always the case that automatically generated product descriptions motivate browsing users to buy. Therefore, we thought it was necessary to evaluate whether automatically generated product descriptions can motivate browsing users to buy.

6.1. *Experiment conditions*

In general, We use BLEU[17]and ROUGE[18] as metrics to evaluate automatically generated sentences by deep learning. However, since these metrics are evaluated by comparing them with the correct answer data, it is difficult to evaluate whether the generated text is a product description that motivates the browsing user to purchase. Therefore, we propose a new viewpoint of the experiment based on Lee et. al. proposed evaluation method[19]. The product description generated by the viewpoints is evaluated. Specifically, the evaluation is performed using the following three viewpoints.

- Broken Axis
  We evaluate whether the generated product descriptions are correct as sentences or not. Specifically, we evaluate whether or not the sentence is grammatically broken, such as repeating the same word repeatedly and not making sense, or breaking off in the middle of the sentence. If the sentence is broken, the rating is 1, and if the sentence is not broken, the rating is 5.

- Correctness Axis
  We evaluate whether the generated product description contains basic information about the product. Specifically, we evaluate whether the product description includes information about the attribute data in the content sheet. If the product information is not

Table 4. Average of subject experiment results.

Method	Broken grade	Correctness	Attractiveness
LSTM	3.54	2.16	1.77
GPT-2	2.62	2.28	1.57
GPT-2 Rinna model	2.80	4.55	3.00

included, the rating is 1, and if a lot of information is included, the rating is 5.

- Attractiveness Axis
  This is the evaluation of the degree to which the user is motivated to purchase the product after reading the generated product description. Specifically, it is an evaluation of whether the user feels the desire to purchase the product after reading the product description. If the browsing user does not feel the desire to purchase the product after reading the product description, the evaluation is set to 1, and if the browsing user feels the desire to purchase the product, the evaluation is set to 5.

When we conducted Experiment 2, we used the same data sets as that used for Experiment 1 (Table 3) Specifically, 10 subjects randomly extracted 40 product descriptions for each method. The total data sets are 120 product descriptions. We randomly presented subjects with a set of 24 product descriptions generated by the three models. The subjects evaluated each axis on a scale of 1 to 5 using the three evaluation axes. The subject repeats this procedure five times for five sets.

6.2. *Results and Discussion*

Table 4 shows the average of the experimentally obtained results.

Broken Axis

The results of the broken axis show that LSTM has the highest average rating of 3.54. The reason for this is that the LSTM consists of a single short sentence, which means that there are few broken sentences, resulting in a high level of brokenness. In addition, the product descriptions generated by GPT-2 and GPT-2 Rinna model sometimes have descriptions in which the assertion changes between the first and second half of the sentence. The reason is that they generate longer product descriptions than LSTM. Table5 shows examples of good results which are rated 4.0 or higher and bad results which is rated 2.0 or lower. The good results of LSTM and GPT-2 are short sentence(s) and straightforward in describing the function of the sofa. The GPT-2 Rinna model is long sentences, but it describes in detail the various functions of the sofa. We can see that these sentences are almost consistent. On the other hand, the bad results for LSTM and GPT-2 both end in the middle of the sentence. The GPT-2 Rinna model looks like a good product description, but if we look closely, it emphasizes disadvantages and the sentences are not a good product description.

Correctness Axis

Next, for the correctness axis, the GPT-2 Rinna model received the highest rating with an average of 4.55. We consider that the reason is that the GPT-2 Rinna model generates descriptions for various items such as product type, color, and material, and combines all of them to include a wealth of basic information about the product. Therefore, the correctness axis of GPT-2 Rinna model accurately reflects the abundance of categories of words used

Table 5. Results of Broken axis

Method	Product Description
Good example: rating 4.0 or higher	
LSTM	The sofa is a cute sofa bed in a two-tone shade of the same color.
GPT-2	The sofa is compact in size. Therefore, it is recommended to match it with a compact system sofa suitable for a studio or single person living.
GPT-2 Rinna model	The color of the sofa is brown or black, so you can choose the one you like best. The sofa is a reclining sofa, so you can enjoy it while lying down and relaxing. Since the sofa has a woodgrain design, you can match it with warm colors such as dark gray, ivory, or brown to tighten the atmosphere of the space and enjoy a different atmosphere! The seat of the sofa is made of urethane, so it is soft and comfortable. This sofa is recommended for families, so it is perfect for those who want a <num>seat, <num>seat, or couch. The material of the sofa surface is synthetic leather, so it is very easy to clean.
Bad example: rating 2.0 or lower	
LSTM	Matching low table:, posted in English
GPT-2	This is a combination of an adult 0-seat sofa, couch, and a compact-sized ottoman.
GPT-2 Rinna model	The color of the sofa is blue, which makes it easy to coordinate with a gentle cafe-style atmosphere. The type of sofa is a corner, so it is recommended for those who are looking for a sofa with <pronoun>like taste. The sofa has a Scandinavian design, so it will probably brighten up the atmosphere of your room. The seat of the sofa is made of cotton, so the disadvantage is that it tends to get stuffy in the summer. Since the sofa is designed for families with children, it is designed to be low enough to keep children occupied on the sofa. You can choose from<eng>& <eng>'s <num>colors. The sofa surface is made of cotton, so it is relatively easy to clean.

in the product descriptions, resulting in a high evaluation. Table6 shows correctness axis examples of good results which is rated 4.0 or higher and bad results which is rated 2.0 or lower. The good results of LSTM and GPT-2 clearly describe the reasons why the product is good. The GPT-2 Rinna model describes why it is attractive, and it also explains in detail the multiple functions of the sofa. The bad results for LSTM and GPT-2 end in the middle of a sentence, as in the broken axis. There are no bad cases in The GPT-2 Rinna model.

Table 6. Results of Correctness axis

Method	Product Description
Good example: rating 4.0 or higher	
LSTM	The color of the sofa changes into a stylish atmosphere because of the green and blue fabric.
GPT-2	The cushion part of the seat is made of urethane, which provides a good cushioning and fit.
GPT-2 Rinna model	The sofa is available in brown or black, so you can choose the one that best suits your room's taste. The <eng>series, which is upholstered in beautiful genuine leather all the way to the back, is designed to give a luxurious yet not oppressive feeling. The type of sofa is a couch, so the width of the seat is a little more generous than <num><eng>, <num>seat sofa. The design of the sofa is Scandinavian, so the sofa itself is also Scandinavian. The seat of the sofa is made of feathers, so it feels nice and smooth all year round. The sofa is designed for a single person, so if you are a family, you may need to change the sofa to a bed. The surface of the sofa is made of genuine leather, so the more you use it, the more it will become attractive.
Bad example: rating 2.0 or lower	
LSTM	High-quality fabric is used for the upholstery, creating a chic atmosphere for
GPT-2	Even though the seat is made of wood and urethane with different heights, the seat section and back.
GPT-2 Rinna model	-

Attractiveness Axis

For the attractiveness axis, the GPT-2 Rinna model obtained the highest rating among the three methods with an average of 3.00. The reason is that the product descriptions generated by LSTM and GPT-2 have many sentence endings in the form of assertions such as " ...is". On the other hand, the product descriptions generated by the GPT-2 Rinna model have sentence ends such as "I recommend it." and "It is attractive". In contrast, the product descriptions generated by the GPT-2 Rinna model had many sentence endings that recommended the product to the recipient, such as "I recommend it. This type of writing style makes browsing users feel that a sentence is attractive. Table7 shows attractiveness axis examples of good results which are rated 4.0 or higher and bad results which are rated 2.0 or lower. The good results of LSTM are short sentences, but it describes the image of the product. GPT-2 describes the image of the sofa in detail. It shows the user how to use the sofa in an easy-to-understand manner. The bad results for LSTM, the grammar is strange and the user did not feel attracted. The GPT-2 is described as elastic and soft, which is a contradiction in terms. The bad results of the attractiveness axis of the GPT-2 Rinna model are the same as a broken result. It means that if the sentence is broken, users do not feel attracted to the product.

Table 7. Results of Attractiveness axis

Method	Product Description
Good example: rating 4.0 or higher	
LSTM	Available in a variety of colors to match your room, enjoy the texture of natural wood.
GPT-2	The cushion part of the sofa seat is made of urethane, which provides a good cushioning and fit.
GPT-2 Rinna model	The sofa is available in black, gray and white, so you can choose the one that best suits your room. It is a couch type sofa, so you can sit comfortably and stretch your legs. The modern design of the sofa gives it a solid a presence that makes it a picture-perfect addition to any living room. The sofa's seat is made of cotton, which absorbs and passes moisture well, making it cool in summer and comfortable even in heated rooms in winter. This sofa is recommended for families with children, so the whole family can get along and make memories together. The material used for the surface of the sofa is cotton, which gives it a light touch.
Bad example: rating 2.0 or lower	
LSTM	The sofa reclining sofa
GPT-2	The seat has a perfect balance of hard polyurethane and high-density polyurethane, making it resilient and soft to the touch.
GPT-2 Rinna model	The color of the sofa is blue, which makes it easy to coordinate with a gentle cafe-style atmosphere. The type of sofa is a corner, so it is recommended for those who are looking for a sofa with <noun>like taste. The sofa has a Scandinavian design, so it will probably brighten up the atmosphere of your room. The seat of the sofa is made of cotton, so the disadvantage is that it tends to get stuffy in the summer. Since the sofa is designed for families with children, it is designed to be low enough to keep children occupied on the sofa. You can choose from<eng><eng><num>colors. The sofa surface is made of cotton, so it is relatively easy to clean.

7. Conclusion

For this study, we have examined three proposed methods, LSTM, GPT-2, and GPT-2 Rinna model as automatic generation methods for product descriptions based on data structures. Specifically, we propose (1) a content sheet consisting of data structures of products, (2) methods for automatic generation of product descriptions based on content sheets, (3) three indicators of correctness, correctness axis, broken axis, and attractive axis to measure the automatically generated product. We conducted experiments of two types. The results of the experiments indicated that GPT-2 Rinna model is best method to generate product description automatically. In the near future, we expect to conduct the following additional studies.

- Rethinking various phrase generation methods for GPT-2 Rinna model sentence generation
- Addition of objective evaluation

- Development of a user interface for an automatic product description generation system.

Acknowledgements

This work was partially supported by Research Institute of Konan University, and by JSPS KAKENHI Great Number 19H04221, 19H04218 and 20K12085.

References

1. S. Hochreiter and J. Schmidhuber(1997), *Long Short-term Memory.*, Neural computation, Vol.9, pp.1735-1780.
2. A. Radford, J. Wu, R. Child, D. Luan, D. Amodel, and I.Sutskever(2018), *Language Models are Unsupervised Multitask Learners.*, Open AI, chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/viewer.html?pdfurl=https%3A%2F%2Fd4mu cfpksywv.cloudfront.net%2Fbetter-language-models%2Flanguage-models.pdf&clen=582775&chu nk=true, 24 pages.
3. J. Cao, J. Gong, and P. Zhang (2019), *Two-Level Model for Table-to-Text Generation*, Proc of the 2019 International Sumposium on Signal Processing System (SSPS 2019), pp. 121–124.
4. H.H. Lee, L.Shu, P. Achananuparp, P.K. Prasetyo, Y. Liu, E. Lim and L.R. Varshney (2020), *RecipeGPT: Generative Pre-training Based Cooking Recipe Generation and Evaluation System*, WWW '20: Companion Proceedings of the Web Conference 2020, pp. 181–184.
5. J. Lee (2020), *Controlling Patent Text Generation by Structural Metadata*, CIKM'20: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pp. 3241–3244.
6. M. Prakhar, D. Chaitali, C. Srinivasa, and S. Srinivasaraghavan (2021), *Automatic Title Generation for Text with Pre-trained Transformer Language Model*, 2021 IEEE 15th International Conference on Semantic Computing (ICSC), pp. 17–24.
7. M. Zhou, J. Zhou, Y. Fu, Z. REn, X. Wang, and H.Xiong (2021), *Description Generation for Points of Interest*, 2021 IEEE 37th International Conference on Data Engineering (ICDE), 6 pages.
8. R. Zhang, J. Guo, Y. Fan, Y. Lan, X. Cheng (2020), *Query Understanding via Intent Description Generation*, Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pp. 1823–1832.
9. T. Zhang, J. Zhang, C. Huo, and W. Ren (2019), *Automatic Generation of Pattern-controlled Product Description in E-commerce.*, In The World Wide Web Conference 2019, pp. 2355–2365.
10. Y. Wang, J. Wang, H. Huang, H. Li, and X. Liu, (2020), *Evolutionary Product Description Generation: A Dynamic Fine-Tuning Approach Leveraging User Click Behavior.*, In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 119–128.
11. S. Kedia, A. Mantha, S. Gupta, S. Guo, and K. Achan (2021). *Generating Rich Product Descriptions for Conversational E-commerce Systems.*, In Companion Proceedings of the Web Conference, pp.349–356.
12. S. Novgorodov, I. Guy, G. Elad, and K. Radinsky (2020), *Descriptions from the Customers: Comparative Analysis of Review-based Product Description Generation Methods.*, In ACM Transactions on Internet Technology, Vol. 20. pp.1–31.
13. Q. Chen, J. Lin, Y. Zhang, H. Yang, J. Zhou, and J. Tang, (2019), *Towards Knowledge-Based Personalized Product Description Generation in E-commerce.*, In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 3040–3050.
14. Y. Taira, S. Sato, R. Miyata, and N. Imagashira, (2019), *Analysis and automatic generation of direct advertising copy.*, In The 25th Association for Natural Language Processing.
15. M. Asahara,(2018), *NWJC2Vec: Word embedding dataset from "NINJAL Web Japanese Corpus".* Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication, pp. 7–25.

16. D.P. Kingma and J. J. Ba, (2015), *Adam: A Method for Stochastic Optimization.*, In International Conference on Learning Representations.
17. K. PapineniS. RoukosT. WardW. Zhu(2002), *BLEU: a Method for Automatic Evaluation of Machine Translation.*, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)pp311–318
18. C. Lin(2004), *ROUGE: A Package for Automatic Evaluation of Summaries*, Proceedings of ACL WOrkshop Text Summarization Branches Outpp74–82
19. C.V. LeeA. GattE. MiltenburgS. WubbenE. Krahmer(2019), *Best practices for the human evaluation of automatically generated text.*, Proceedings of The 12th International Conference on Natural Language Generationpp. 355-368