

Improvement of Story-telling Advertisement According to Screenwriting Techniques

Daiki Uehara

Graduation School of Information Science and Engineering
Ritsumeikan University
1-1-1 Nogihigashi, Kusatsu, Shiga, 525-8577, Japan
Email: d.uehara5151@de.is.ritsumeai.ac.jp

Fumiko Harada, Hiromitsu Shimakawa

Connect Dot Ltd.
160-2 Saihojicho, Nakagyo-ku, Kyoto, Kyoto, 604-0866, Japan
Graduation School of Information Science and Engineering
Ritsumeikan University
1-1-1 Nogihigashi, Kusatsu, Shiga, 525-8577, Japan
Email: {harada, simakawa}@cs.ritsumeai.ac.jp

Abstract—The study proposes a method for enterprises to enhance their scenarios for storytelling marketing. Scenarios to attain the sympathy of consumers are hard to write for enterprisers unfamiliar with scenario writing. In this research, we focus on the story development and scene development of the screenwriting technique. This method uses logistic regression to detect expressions lacking in each scene of the scenario. Repeated modification of the scenario according to the indication of the lacked expression leads to scenarios to an advertising document that resonates with the reader. In this study, we verified the effectiveness of the method for specific story development. As a result, it turns out the proposed method can support the creation of scenarios whose scene development is close to the model ones. The method enables enterprisers to refine the scenario into advertising documents that resonate with readers.

I. INTRODUCTION

THE SPREAD of smartphones greatly increases our opportunities to view advertisements on the Internet. Smartphones promote privacy and immerse browsing on advertisements anytime and anywhere without worrying about what others are thinking.

Information on products and services enterprisers want to sell on the Web such as SNS that is called Social Network Services and reputation sites play an important role in their marketing [1]. Positive impressions from actual users of the products and services reduce skepticism of new customers against them, which makes it easier for the customers to get interested in them. It implies the advertisements should include the user's experiences of using the products and services. When enterprisers make advertisements, they should include experiences and impressions of customers actually using the products and services. Incorporation of the customer's view into the advertisement enables the reader of the advertisement to have a simulated experience of using the products and services.

On the other hand, in order for the readers to have simulated experiences with the advertisement, the advertisement scenario needs to attract the reader's interest. Advertisements on SNS and other Web media have been shown to have positive effects on consumers' motivation to buy. Especially, the storytelling marketing method has been away attracting attention to get consumers' empathy using narrative advertising [2] [3]. Its

most prominent feature is that readers of the advertisement feel empathy with products and services through a scenario telling efforts of enterprisers who invented them. According to the study by Laer [4], commercial advertisements should use a narrative format to reduce customers' resistance to buying products and services. However, it is almost impossible for enterprisers who have no experience to write scenarios attracting the attention of consumers. On the other hand, a huge cost is necessary to commission professional scenario writers to make an advertisement. Fortunately, there is a book summarizing how to write scenarios. The study refers to a set of know-how explained there as screenwriting techniques. This study proposes a method to support writing narrative scenarios from the enterprisers' actual experiences based on the screenwriting technique. The purpose of screenwriting is to write a scenario through which the main character can get the reader's empathy. Advertisements based on a scenario attaining empathy from readers are posted on SNS so that it would show the enterprisers' efforts in their commercialization. It is expected to get the customers' empathy. If enterprisers follow screenwriting techniques to use appropriate expressions, they can write a scenario in which the reader can empathize with the business. Empathy leads to turning the reader into a customer.

The study focuses that scenarios should use words characteristic to each scene. A lack of words suitable for each scene would prevent readers from getting empathy. The development of scenes varies with a storyline. The proposed method creates a logistic regression model that classifies the scenes of a scenario for a specific storyline. The enterprisers select a storyline that fits their actual experiences to write the scenario for each scene. If there is a lack of the scene words in each scene scenario, the logistic regression model could not classify the scenes correctly. The proposed method detects the lack of expressions for the scenes which are not correctly classified. Through the detection of the lack of expressions, the method can support the enterprisers to modify scenarios that get empathy from readers.

In the experiment, a logistic regression model is created to classify scenes for a specific storyline. From the results of the experiment, the paper discusses how to point out the lack of the scene words. By evaluating the appropriateness of

words characteristic to each scene in the particular storyline, the method has turned out to have the potential to be extended to other storylines. As the enterprisers improve their scenarios, they can improve their scenarios to get the reader's empathy.

The scenario improvement by the enterprisers themselves will significantly reduce advertising costs.

If enterprisers follow the screenwriting techniques, they can write scenarios that can get the readers' empathy even without professional knowledge about scenario writing. This method enables even enterprisers of small businesses to create advertising documents which get the reader's empathy. These advertisements will contribute to business expansion.

II. CREATING A DOCUMENT TO GET THE READER'S EMPATHY

A. Importance of Empathy in Marketing

Painful experiences have the property of being easily transmitted to others. Understanding and sharing others' experiences full of effort, we get easier to empathize with them.

There is a marketing method called storytelling marketing, which uses narrative advertising to get the sympathy of consumers. Empathy in storytelling marketing is that the consumer gets emotionally involved with characters in the advertisement. The hard business experience of enterprisers is an experience full of efforts. The enterprisers can make advertisements that are easy to get the sympathy of readers by sending the stories of their experiences full of efforts as stories. Using narrative advertising makes it easier for the consumers to feel sympathy for the characters in the advertisement.

In marketing, it is important to get the empathy of customers. We would easily empathize with hard experiences where someone has commercialized products and services. For this reason, storytelling marketing often uses scenarios that tell the story of the enterprisers' own hard experiences as advertising documents. Empathy toward narrative advertising increases consumers' interest in the product and motivates them to purchase it.

B. Enterprisers with Poor Documentation Knowledge

There are many ways to create narrative advertising documents. Many of these documents are created from scratch. Enterprisers who write narrative advertising documents need to carefully consider the structure of the stories. However, most of the enterprisers who write narrative advertisements have little or no experience in writing stories. For enterprisers unfamiliar with writing scenarios, it is difficult to come up with a narrative structure specific to a storyline.

There are methods that can automatically generate new stories by learning several stories. The method takes the context into account to generate a story. However, the context of each sentence is incoherent because the words and storyline are taken from various kinds of stories. Incoherent sentences make it difficult for readers to understand the content. Such sentences fail to get empathy from the readers because the readers cannot understand the content of the user's document. Throughout the documentation, enterprisers themselves are

required to write the contents so that almost all readers can understand it.

By writing the advertising documents for themselves, enterprisers can avoid losing the excitement and authenticity in their actual experience. The emotional changes of the characters are important. The enterprisers are characters in a narrative advertising document. In order to directly convey the enterprisers' hard experiences to readers, the enterprisers themselves should write the narrative advertisement documents. On the other hand, it is difficult for enterprisers themselves to evaluate the completeness of their documents. In many studies, the evaluation of the produced document depends on the reader's feelings. Since each reader's feelings are different, it is not appropriate to use feelings to evaluate a story. Enterprisers should be able to objectively evaluate whether the advertising document gets the readers' empathy. To write a scenario for storytelling marketing that appeals to many readers, we need a method satisfying the following two requirements. First, the method should help enterprisers to create documents with emotional changes so that the readers can understand their hard experiences. Second, the method should enable enterprisers themselves to objectively evaluate the completeness of written documents.

C. Screenwriting Techniques

Let enterprisers write narrative scenarios for storytelling marketing to get readers' empathy from their own actual experiences.

Snyder has systematically organized the many famous Hollywood movies to summarize their scenario creation methods into the book, *Save-the-cat* [5]. This paper refers to the summary as screenwriting. The purpose of screenwriting techniques is to present secrets to writing a story attaining the reader's empathy for the main character. Screenwriting techniques help the enterprisers to write scenarios from their actual experiences to get the readers' empathy.

Snyder examined the narrative structure of many famous Hollywood movies to find the movies have commonalities. The commonalities are a combination of the small number of patterns relating to the overall flow of the storyline and a single pattern relating to the development of the scene in which the characters' experiences happen. The screenwriting techniques by Snyder present the combination of these two patterns. This paper refers to scenarios by experts according to the combination of these two patterns as pro-scenarios.

The pro-scenarios would be compliant with one of 10 storylines. The 10 storylines are shown in TABLE I. The left column specifies the name of the storyline. The content of each story is written in the right column.

The structure of many famous Hollywood movies has turned out that the characters' experiences are expressed by 15 scenes. TABLE II shows the development of the 15 scenes. The first column shows the names of the 15 scenes, while each of them is explained in the second. All scenarios exemplified in the screenwriting techniques use this scene development to represent scenes and their development.

TABLE I
STORYLINE;

COURTESY OF BLAKE SNYDER, SAVE THE CAT!: THE LAST BOOK ON SCREENWRITING YOU'LL EVER NEED, MICHAEL WIESE PRODUCTIONS, 2005

Storylines	Story
Monster in the House	The main character' escape from a monster that appers in a closed environment.
Golden Fleece	The main character who goes on a journey and obtains something important other than what he was initially seeking.
Out of the Bottle	The main character who obtains a mysterious or great power, goes through various experiences, and finally accomplishes something without relying on that power.
Dude with a Problem	The ordinary main character who gets caught up in an extraordinary event.
Rites of Passage	The main character stands at a crossroads in his life, and after suffering, he accepts his true self.
Buddy Love	The main characters are lacking in something, but they make up for their shortcomings and grow together.
Whydunit	When the essence of the story begins to emerge, the view that had been visible changes completely.
The Fool Triumphant	The main character who is thought to be dumb and incompetent, but unexpectedly triumphs.
Institutionalised	The struggles of the main character who lives in a special family and organization.
Superhero	The main character has a special power and fate, but he suffers because he is special.

TABLE II

AN EXAMPLE OF SCENE DEVELOPMENT;

COURTESY OF BLAKE SNYDER, SAVE THE CAT!: THE LAST BOOK ON SCREENWRITING YOU'LL EVER NEED, MICHAEL WIESE PRODUCTIONS, 2005

Scene	Context
1. Opening Image	Sets the tone, mood and style. Often introduces the main character and their "before" state.
2. Theme State	Someone will pose a question or make a statement (usually to the main character) that is the theme of the movie, the thematic premise. An argument stated, and the rest of the film is the argument laid out, proving or disproving the statement.
3. Set-up	Introduce all the main characters. Introduce every character behaviour that will need to be addressed later on, that will need to change if the hero is to win. Introduce the things the hero needs, or need fixing, or are missing from their life (SHOW them). The world before the adventure starts. The thesis.
4. Catalyst	Something that arrives, a message, an event, that changes things. Bad news, but by the time the adventure is over, it's what leads the hero to happiness. The first moment when something happens.
5. Debate	Hero thinks this is crazy. Should they go? Is it possible? How can they do it? It should ask a question of some kind.
6. Break into two	We leave the old world behind, into the antithesis. No later than page 25 (of a 110 page script). The hero must make the decision themselves to step into Act Two.
7. B Story	Carries the theme of the movie. Often "the love story". Often a bunch of entirely new characters, maybe opposites to those from Act One.
8. Fun and Games	Provides "the promise of the premise". Not as concerned with forward progress. Lighter than the rest of the movie. Where many of the trailer's moments are found. Set pieces. A break from the stakes of the story.
9. Midpoint	For the hero it seems like all is won or all is lost. Fun and games are over, back to the story.
10. Bad guys close in	Hardest to write. If all seemed won at the midpoint, now things start to go wrong. Dissent, doubt etc. disintegrate the hero's team and the defeated bad guys/thing regroup and return. There is nowhere for the hero to go.
11. All is lost	The opposite of the midpoint in terms of an "up" or a "down". A false defeat. It might seem like a total defeat. It can help to have something about death here. e.g. someone/something dies or thinks about death.
12. Dark night of the soul	Hero is hopeless, at their lowest point, no one to help them, no ideas. The solution is found. Thanks to characters in the B story, the conversations about the theme in the B story, the hero trying to find ways to beat the bad guys in the A story.
13. Break into three	The two stories meet and intertwine.
14. Finale	Act three. Lessons learned are applied, character ties are mastered. A and B stories end in triumph for the hero, the bad guys are dispatched (in ascending order), the old world is turned over, a new world is created.
15. final Image	The opposite of the opening image. Proof that change has occurred.

All storylines adopt the development of the 15 scenes. However, the contents to be stated in the scenes varies with each storyline. Therefore, the scene developments applicable to all storylines are too abstract for enterprisers to use as a reference for writing scenarios. When enterprisers write narrative scenarios for storytelling marketing, the scene development embodied for each storyline is preferable, because it gives enterprisers more specific information on each scene.

III. IMPROVEMENT OF ENTERPRISER SCENARIO

A. Improvement toward Pro-scenario

When a beginner tries to write a script, it is popular for the beginner to imitate the script of an expert in the way the expert writes the script. The beginner will progressively improve his script, increasing its similarity to the script of the expert. In a situation where the expert is unable to accompany the beginner for instruction, the following two methods are necessary. The first objectively evaluates the similarity between the beginner's script and the expert's script. The second makes it clear what parts of the beginner's script are not similar to the expert's script.

The proposed method evaluates the similarity between a scenario by an enterpriser and a pro-scenario using text analysis techniques. Snyder presents 10 kinds of storylines. The pro-scenario has detailed scene development for each storyline. The enterpriser selects one storyline that applies to their actual experience. After a choice of a storyline, the enterpriser is informed of the content of each storyline and its application example to the actual life. Since the 10 storylines are only rough explanations of scene development in famous fascinating stories, it is hard to find out which storyline is suitable to represent the actual experiences of the enterpriser. Therefore, the proposed method provides the enterpriser with examples of the actual events in the actual life in each storyline. An example of a "monster in the house" is the business management crisis caused by the COVID19 pandemic. The enterpriser can easily choose a storyline, comparing his actual experience with the examples. After the choice of the storyline, the enterpriser starts to write the scenario of the actual experience according to the scene development in the selected storyline.

The scenario for advertising is assumed to be for SNS and blogs. The total number of words in the scenario is much smaller than in the ones for movies. The division into 15 scenes in screenwriting is excessive for a scenario with few words. To accommodate advertisements in SNS and blogs, the number of scenes is reduced from 15 to 4, based on scene switching. The method takes it after the model of the four-part organization of Chinese poetry. The enterprisers write their actual experiences in the form of a scenario that follows the scene development corresponding to the chosen storyline.

There are five examples of movies for each of the 10 storylines, which are used as references to explain the screenwriting technique. The scene development of these movies is explained by presenting a scenario in each of the 15 scenes. In each of the movies, the scenario is presented for each scene of the movie. The scenario of the enterprisers according to the screenwriting technique should use expressions similar to these pro-scenarios. It makes the scenarios of the enterprisers more empathetic to readers.

Based on the similarity between the scenario written by the enterprisers and the pro-scenario, it is possible to improve the scenario of the enterprisers by giving specific examples of the pro-scenario to the enterprisers to improve their scenario.

According to this idea, the scenario of the enterprisers is improved by the following process.

- 1) The method finds which storyline the enterprisers' scenario is similar to.
- 2) The method shows the enterprisers some examples of the scenarios based on the similarity of storylines.
- 3) The method let the enterprisers improve their scenarios, comparing the given examples with their own scenarios and confirming the lack of expressions in the scenarios.

The proposed method identifies the deficient expressions in the enterprisers' scenario from the features of each scene. It also presents specific correction suggestions.

B. Improving the Scenario by Focusing on Words

In this study, the proposed method improves scenarios by detecting missing expressions from actual life scenarios described by employers according to screenwriting techniques. Fig. 1 shows the outline of this method.

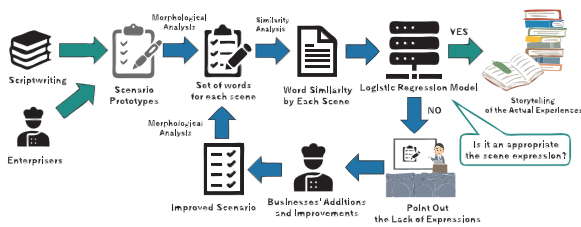


Fig. 1. Method

A screenwriting scenario has 15 scenes. Each scene has its content to be described. The enterprisers create the scenario according to the content. In addition, there are keywords for each scene, and these keywords represent the characteristics of the scene. This paper refers to this keyword as the scene word. By using appropriate scene words, the enterprisers can create a scenario that will gain the reader's empathy.

There are multiple words used depending on the context. Enterprisers do not necessarily need to use the scene words. The proposed method regards the words with meaning similar to that of the scene words as an appropriate expression. The word is defined as a synonym of the scene word. This method evaluates the scenario written by the enterprisers by the degree of similarity to the pro-scenario.

In this study, the scenario written by the business owner is defined as the target scenario. Morphological analysis is used to extract nouns, verbs, adjectives, and adverbs for each scene in the target scenario. From the morphologically analyzed target scenario, the number of synonyms of scene words can be found for each scene. In the same way, from the morphologically analyzed pro-scenario, the number of synonyms of scene words can be counted for each scene. The number of synonyms of scene words is used to detect the similarity of each scenario.

A logistic regression model is used to detect similarity. The model is trained to classify the scenes in the target scenario

by the number of synonyms of the scene words in the pro-scenario.

When each scene in the target scenario is more correctly classified by the logistic regression model, the similarity with the pro-scenario is high. If there is a high degree of similarity between the scenarios, the target scenario can be concluded to be an appropriate representation of each scene. If the similarity between the scenarios is low, the lack of expressions in each scene is pointed out to the enterprisers. The lack of expressions is identified from the scene words in each scene. Based on the suggestions, the enterprisers modify the target scenario. The modified target scenario is analyzed again by morphological analysis, and the scene words are extracted for each scene. The modified scenarios are classified using a logistic regression model to detect similarities with the pro-scenario. In the repeated suggestion of the missing expressions which triggers modification of scenarios, the enterprisers improve the target scenario. By extracting the scene words in the pro-scenario, the enterprisers can find out what words should be placed in each scene.

C. Documentation of Actual Experiences

In marketing, it is important to create scenarios for advertisements to gain the empathy of readers. The book, *Save-the-cat*, presents screenwriting techniques to create scenarios to make readers feel empathy. This study values the 10 storylines and 15 scenes of the screenwriting techniques.

First, enterprisers are requested to select a storyline that matches their experience from the 10 storylines. Since the content the enterprisers write in each scene varies with each storyline, the enterprisers need to select a storyline at the beginning. It is difficult for enterprisers to have to decide for themselves which storyline best fits their experience. To help them, some explanations of the storyline development should be provided to the enterprisers in the study.

Second, the enterprisers create a scenario to fit the content of the 15 scene developments in the selected story. In this study, the fifteen scenes are classified into four following a storytelling style. When creating a scenario, the method provides enterprisers with explanations of the scene development. The explanations of each scene promote the enterprisers to create the scenario smoothly.

The scene words to be used for each scene are presented in the explanation of the scene development. For example, the first scene, which sets up the story, should state many places, times, and proper nouns relevant to the actual experiences. The scenario that follows the scene development should have a representation similar to the pro-scenario. The choice of storylines suitable for their actual experiences along with the description of scenarios compliant with the scene development would enable enterprisers to create scenarios that make readers empathize. Understandably, it is difficult for enterprisers to create a scenario similar to professional scenarios with only a single trial. Therefore, the proposed method repeats the process where the enterprisers notified of the lack of expressions in each scene of the target scenario modifies the scenario by

themselves. Through the processes, the enterprisers gradually improve their target scenarios.

D. Finding Lack of Expressions in Each Scene

Enterprisers aiming at creating scenarios similar to the professional scenario need indications telling what parts to be modified. The proposed method points out the lack of expressions, depending on the number of synonyms of the scene words in the target scenario.

Logistic regression is used to find the lack of expressions from the similarity between the target scenarios and the pro-scenarios. The logistic regression model is trained so that it may predict which of the 4 scenes the current one belongs to, using the number of appearances of synonyms of the scene words. Let the logistic regression model classify the scenes in the target scenario. The model is expected to classify the scenes of the target scenario into ones intended by the enterprisers. The classification which is out of expectation implies the scene written by the enterprisers lacks proper expression.

The words to be learned by the model are taken from the pro-scenarios. For example, suppose the scene word for the first scene is "strange". The method counts up synonyms for "strange" in the first scene of the pro-scenario. It also counts up synonyms for "strange" in the remaining scenes. Since "strange" is involved in the scene words in the first scene, its synonyms should appear most in the first scene. In the same way, the number of synonyms for each scene word in the other scenes will be used to identify the scene. The model can classify scenes through the training of a logistic regression model with the number of occurrences of synonyms of scene words in each scene as explanatory variables. It also contributes to finding the lack of expressions.

The logistic regression model gives the classification results as probabilities. If the scene expression written in the enterprisers' scenario is not appropriate, the probability is high for an undesired scene.

The probabilities show the expressions that should be deleted and those that are lacked in the enterprisers' scenario. If the scenes in enterprisers' scenarios are not classified into appropriate ones, it is likely to contain the scene words from other scenes. Those words should be removed. At the same time, to make each scene of the enterprisers' scenarios classified appropriately, the enterprisers should add synonyms of the appropriate scene words to the scenarios. Since the number of synonyms is used as an explanatory variable, the model can calculate the occurrence ratio of the synonyms in each scene. Based on the ratio, the proposed method can suggest to the enterprisers what kinds of expressions lack in the scenes in their scenario. By specifically pointing out the lack of expressions in the scenes of the enterprisers' scenario, the method can support the enterprisers to modify the scenario by themselves.

IV. EXPERIMENT

A logistic regression model is created from the pro-scenario. Since the study aims at the improvement of target scenarios

written in Japanese, it is assumed that both target scenarios and pro-scenarios are written in Japanese. From the scene words of the pro-scenario, a model is created to classify the four scenes of each storyline.

A. Purpose of the Experiment

In this study, the scene words are used to divide the story into four scenes. Since the scene words depend on the storylines, it will take a large amount of time and effort to check the effectiveness of this method for all storylines. To confirm the effectiveness of this method, it is necessary to create a model to classify the scenes for a specific storyline of screenwriting.

In this experiment, the storyline of "Monster in the House" is used to create the model. The "Monster in the House" scenario has five scenarios written by professional writers. The five scenarios are the stories of five movies: "Alien", "A Dangerous Affair", "Scream", "Saw", and "The Ring".

First, the text expressing the story is divided into some words by morphological analysis. For morphological analysis, Mecab's -Ochasen is used. By reading the five scenarios for each of the four scenes divided according to the screenwriting technique, it is possible to find out the scene words for each of the four scenes. A logistic regression model is created by counting the number of synonyms of scene words in each of the five scenarios, using the number of scene words as the explanatory variable and the scene as the objective variable. The logistic regression model created classifies the four scenes of "Monster in the House". Analyzing the appropriateness of the scene words, it is necessary to check the coefficients of the scene words for each scene in the logistic regression model.

B. Word Extraction for a Specific Storyline

A logistic regression model is used to classify the scenes in the target scenario. The synonyms of the scene words are used as explanatory variables in the logistic regression model.

By reading the five pro-scenarios for each storyline, it is possible to identify the common expressions in each scene as scene words. Based on the similarity between the scene words and the words in the scenario, the number of times synonyms of the scene words appear is counted. That number of times synonyms is used as the feature value of each scene.

The fasttext was used to generate the word vectors. On the other hand, the scene words do not have to be words that appear in the scenario. This is because using fasttext does not necessarily mean that the scene words in the pro-scenarios accurately capture the meaning of the scene. To detect scene words in a scenario, computing the Euclidean distance between the multiple scene words and all words in the scenario is necessary. When the synonym of the scene words appears among words with close Euclidean distance, these scene words are appropriate.

The number of scene words and synonyms for each scene is shown in TABLE III to TABLE VI. The words in the top two columns of the table are scene words, and the column on the far left is the examples of the movies. The scene words used in

each scene are as follows. The scene words in the first scene are the number of nouns related to "place" and "time." The scene words in the second scene are a number of synonyms for "strange," "anxious," and "misbehave." The scene words in the third scene are the number of synonyms for " progressively " and "notice." The scene words in the fourth scene are the number of synonyms for "finally" and "courageously."

TABLE III
NUMBER OF SCENE WORDS AND THEIR SYNONYMS IN THE FIRST SCENE

movie	place	time	strange	anxious	misbehave	progressively	notice	finally	courageously
Alien	7	1	0	0	0	0	2	0	1
Danger	5	0	0	0	0	0	1	0	1
Ring	4	2	0	0	1	0	1	0	0
Scream	2	0	1	0	0	0	0	0	0
Saw	5	1	2	0	0	1	2	0	0

TABLE IV
NUMBER OF SCENE WORDS AND THEIR SYNONYMS IN THE SECOND SCENE

movie	place	time	strange	anxious	misbehave	progressively	notice	finally	courageously
Alien	3	0	6	1	2	2	0	0	0
Danger	5	1	0	4	2	1	0	0	0
Ring	1	1	5	0	1	3	6	0	0
Scream	5	1	1	4	2	1	4	4	2
Saw	3	1	1	0	0	1	9	0	1

TABLE V
NUMBER OF SCENE WORDS AND THEIR SYNONYMS IN THE THIRD SCENE

movie	place	time	strange	anxious	misbehave	progressively	notice	finally	courageously
Alien	0	0	0	3	0	3	4	1	0
Dangerous	4	1	0	2	0	3	3	1	0
Ring	3	2	4	0	2	4	4	0	1
Scream	3	0	0	2	0	2	0	0	0
Saw	2	0	1	0	0	0	8	1	0

TABLE VI
NUMBER OF SCENE WORDS AND THEIR SYNONYMS IN THE FOURTH SCENE

movie	place	time	strange	anxious	misbehave	progressively	notice	finally	courageously
Alien	1	0	0	0	1	0	1	2	4
Danger	4	0	0	0	0	0	3	0	1
Ring	3	0	0	0	2	0	4	0	1
Scream	2	0	0	0	0	0	2	0	4
Saw	5	0	1	1	0	0	1	1	4

In scenes 2 through 4, the number of synonyms of the scene words are taken as the explanatory variable. On the other hand, in the first scene, the number of words representing "place" and "time" is the feature value. The first scene is where the words related to the setting of the story scene or time should appear. Since each story has a different way of expressing "place" and "time", it is difficult to find the synonyms of scene words in the first scene. In the first scene, there should be many nouns that indicate "place" and "time". For this reason, only in the first scene, the noun types were referred to from the category information of Janome, and the number of these nouns was used as an explanatory variable. Morphological analysis of nouns in Janome can get the information of the category of the noun. The categorical information is given in the form of noun types, such as common and proper nouns, as well as personal names and places. The number of nouns that are "place, the organization" and "time" among the category information was used as the explanatory variable in the first scene.

The first scene is the scene that sets up the story, but the nouns related to people's names are not considered as features. The nouns about people also appear in other situations. Therefore, it is difficult to use them as features only for the first scene.

A logistic regression model is trained using the number of synonyms of the scene words as a feature. The appropriateness of the scene words for each scene is evaluated by the scale of the coefficients of the logistic regression. For the first scene, if the coefficients of "place" and "time" are large, the scene word is appropriate. The same method is used for scenes 2 through 4 to evaluate the appropriateness of the scene words.

C. Creating the Scene Classification Model

To find the synonyms of the scene words, the similarity between the scene words and all words in the pro-scenario is calculated. There are two main types of similarity calculations: Euclidean distance and cosine similarity. For the reasons given below, this study adopts the Euclidean distance. The cosine similarity between words with similar meanings should be close to 1.0. For example, the cosine similarity between "home" and "father" should be close to 1.0. However, the actual cosine similarity was 0.431. In fasttext, it was determined that "household" and "father" are not words that are close in meaning. On the other hand, the cosine similarity may be expected. As shown above, when using fasttext to vectorize words, it is difficult to uniquely define the threshold with cosine similarity.

In some cases, the word meanings that the fasttext for generating word vectors is trained with different word meanings expected. As a result, it is necessary to manually set an appropriate Euclidean distance threshold. When the threshold was 3.7, the most appropriate synonyms for the scene words were found. The number of words with Euclidean distance less than 3.7 was used as the features.

When the coefficient of the scene word set for each scene is large, the scene word is appropriate because the selected scene word is needed to classify each scene. When the scene word is appropriate, the threshold set to count synonyms of the scene word is also appropriate.

D. Word Vector

The delay in the response of a method to the user's input should be as small as possible. To reduce the delay in the response from the method, the processing in the method should be faster. In this method, fasttext is used to vectorize the words in the scenario. fasttext is fast computing large dimensional vectors such as natural language.

Although fasttext used in this study is trained on many Japanese words, it does not cover all of them. The words that have not been learned by fasttext are called undefined words. There are many undefined words in the scenario. In many situations in natural language processing, undefined words are ignored. Nevertheless, in our method, the undefined words may be in the scenario. Therefore, it is not possible to ignore the undefined words.

In this study, fasttext model is used by applying the python package "Magnitude" to fasttext, so that all words in a scenario may be vectorized. Magnitude gives a vector of unknown words in the fasttext that are similar to the undefined word in terms of letter order and the synonyms of the word. Magnitude considers the words with similar letter sequences to be close in meaning. By referring to vectors of the synonyms of the words with similar letter sequences to the undefined words, the meaning of the undefined words is more grounded.

E. Scene Word Selection for Each Scene

In the study, scene words are selected subjectively. Their effectiveness is confirmed with a logistic regression model.

In the five scenarios of "The Monster in the House," the common expressions and words appear in each scene. The story of "The Monster in the House" is about the main characters' survival from a monster that appears in a closed space. Monsters are not only monsters but can be people or disasters. Monsters can be seen as things that worsen the main character's situation. A closed space is not only a building, but also a region, a country, or even space if it is difficult for the main character and people around him to escape from the monster. Organizations, such as families, are also spaces.

The first scene of "The Monster in the House" introduced the characters and gave information about the place and time. The scene words in the first scene are the words related to "place, the organization" and "time. In the second scene, the protagonists face a monster. Not accidentally, the main characters have made a blunder that causes them to face the monster. However, the main character hasn't noticed his blunder yet. Therefore, in the second scene, the main character just feels unsafe. The characters feel strange about the situation and a little unsafe about their behavior. As a result, the three scene words for the second scene are "strange," "anxious," and "misbehave." In the third scene, the main character notices his "failure" and the situation around him progressively worsens. The main character is forced into an irreversible situation. In the third scene, there are two scene words, "gradually" and "notice". In the fourth scene, the main character forced recovers and overcomes the situation finally. There was a lot of content about the main character using his powers to escape from the monsters courageously. The two scene words in the fourth scene are "finally" and "courageously".

The scene words in each of the four scenes were selected subjectively. The appropriateness of these scene words is tested by the coefficients of a logistic regression model. When the coefficient of the scene word selected in each scene is large, the scene word is appropriate.

F. Experimental Results

From the scene words, a logistic regression model is created to classify the scenes in the target scenario. Table VII shows the results for the coefficients of the logistic regression model.

In the second scene, the coefficients of "strange" and "anxious" were 0.703 and 1.023, respectively, which were larger than those of the other scene words. The words "strange"

TABLE VII
REGRESSION COEFFICIENT OF SCENE WORDS

	place	time	strange	anxious	misbehave	progressively	notice	finally	courageously
scene1	0.432	0.336	-0.012	-0.78	-0.335	-0.427	-0.605	-0.1	-0.413
scene2	0.078	0.075	0.703	1.024	-0.209	-0.109	0.458	-0.18	-0.056
scene3	-0.41	-0.021	-0.301	0.044	-0.133	0.912	0.145	0.357	-0.359
scene4	-0.101	-0.39	-0.39	-0.288	0.258	-0.376	0.001	-0.078	0.828

and "anxious" are appropriate as scene words in the second scene. This result indicates that the second scene in which a disturbing atmosphere is generated can be represented.

On the other hand, the expression "misbehave" is a numerical value close to the coefficients in other situations. The third scene is the best scene word because the coefficient of "progressively" is 0.912 which is a large value. In other words, the expression "progressively increasing fear" is unique to the third scene.

The fourth scene is appropriate as a scene word because the coefficient of "courageously" is 0.827 which is a large value. However, since the coefficient of "finally" is smaller than that of the scene words in other scenes, the expression is not unique to the fourth scene.

The coefficients of the subjectively selected scene words are larger for the second through fourth scenes. In the first scene, the coefficients of "place" and "time" are smaller than those of the other scene words. The words related to "place" and "time" do not show the characteristics of the first scene. On the other hand, the coefficients of "anxiety" and "notice" are larger. Table VII shows that the number of synonyms for "anxious" and "notice" in the first scene is lower than in the other scenes. In other words, when classifying the first scene, there must be few synonyms of "anxious" and "notice". The first scene should be written with many words about the setting of the story. In addition, the words related to emotions and actions should be written less frequently.

It is possible to create a model for classifying scenes based on the words and expressions specific to each scene in the movie scenarios.

V. WORDS TO DESCRIBE SCENE DEVELOPMENT

A. The Importance of Words to Describe a Situation

The coefficients of the logistic regression model indicate that it is possible to classify scenes by scene words. In screenwriting, the story is developed into 15 scenes. In this study, the 15 scenes are divided into four categories.

The main character's emotions are considered to be a major factor in the development of the scene. Scenes in screenwriting are developed through the emotional ups and downs of the main character. The main character's emotions change in the following order: positive, negative, positive. The words that express emotion, such as "happy" and "sad," are called emotion words. Based on the assumption that the emotional ups and downs of the protagonist were responsible for the scene development, the choice of emotional words should have been important for the scene words. The scenes of the target scenario are classified according to the number of synonyms of the scene words appearing in the pro-scenario.

The scenes of the target scenario are classified according to the number of synonyms of the scene words appearing in the pro-scenario. Words that are not common to all pro-scenarios cannot be said to capture the characteristics of the scene. Therefore, the synonyms of scene words must occur in all pro-scenarios. Emotional words are not appropriate as common scene words because they do not appear often in pro-scenarios.

In movies and novels, emotions are often expressed through the facial expressions and actions of characters. In addition, the emotions of the characters may be expressed by the surrounding circumstances. For this reason, emotional words are rarely written in professional scenarios. For example, assume a situation where the main character is in a crisis. In novels and movies, the main character's sense of danger is conveyed to the reader through the main character's pained expression and the description of a situation from which there is no escape. Even if the main character is verbalizing his emotions, facial expressions and the situation around him should be described. The description of the main character acting hard gives the readers a detailed picture of the main character's situation. The adverbs that modify the main character's actions are also important in conveying the main character's emotions.

The enterprisers create a target scenario with a panoramic view of the actual experience. The expressions of the main character, the enterprisers themselves, and the situation around them must be described from a panoramic view. In the target scenario, it is important to describe the main character's facial expressions, surroundings, and adverbs. Scene words should be chosen to describe the surrounding situation or adverbs that describe the protagonist's actions rather than emotional words.

B. Meaning of the Words in the Story

The words selected as the scene words are based on fasttext. Fasttext can vectorize words and output synonyms of the vectorized words. In this study, synonyms of scene words were checked by fasttext, and appropriate scene words were selected subjectively.

Note that fasttext uses Facebook as its corpus. Due to it, discrepancies may happen in the meanings of words used in scenarios. For example, the word "place" is commonly used to mean "place" in English. However, the top 30 synonyms for "place" in fasttext were sumo-related words such as "yokozuna". This is probably because there were many posts about sumo on Facebook when fasttext model was learning the word "place". In "The Monster in the House," the first scene is characterized by "place," meaning "place". In our method, "place" was not selected as the scene word, but the number of nouns representing a place was used as a feature. This is because when "place" is used as the scene word, the similarity with the word meaning "place" is very low.

Due to the use of fasttext in our method, there are words such as "place" which have a different meaning in the scenario. Currently, the words selected as scene words must be checked in the fasttext to ensure that they have the appropriate meaning for the scenario. One of the ways to solve this problem is to

actually create a corpus with many movie scripts and learn the model of fasttext. Learning from the movie script is also important for creating vectors with Magnitude. Magnitude generates vectors from words and their synonyms with a similar word-letter sequence to the undefined word. If the synonym of a word with a similar letter sequence to the unknown word is a word with an unsuitable meaning for the scenario, the meaning of the word will be different from the original meaning of the undefined word. It is important to learn fasttext in movie scripts in order for unidentified words to have proper meaning.

C. Expandability to Other Storylines

In the proposed method, when scenes in the target scenario are not properly classified, the lack expressions are identified based on the difference in the number of synonyms of scene words between the pro-scenario and the target scenario. A specific model was created for "monsters in the house," and the appropriateness of the scene words was tested by the size of the coefficient of determination.

The coefficients for the scene words "strange" and "anxiety" in the second scene were larger. This indicates that sentences containing "strange" or "anxiety" are classified as the second scene. In other words, by setting a scene word for each scene, the model is created to classify each scene.

The scene words from the pro-scenario were used to create the classification model. By using scene words, it is possible to identify the lack of expressions in the target scenario. The scene words enable the enterprisers to describe the scene appropriately. The presence or absence of the scene words in each scene can be used to identify the missing expressions in the target scenario. The scene words are considered to be present in every storyline. Therefore, for the other nine storylines, the model for the scene classification can be created by selecting the scene words for each scene and counting its synonyms.

VI. APPROACHES USING OTHER THAN SCENE WORDS

A. Vectorization of Words and Documents

To calculate the similarity between texts, it is general to vectorize words or documents. For vectorization of natural language, word2vec is generally used. Word2vec vectors the documents from the number of appearances of words. However, the generated document vector depends on the number of appearances of the words. Therefore, word2vec ignores the context. Ignoring context makes it difficult to detect whether the scenario is appropriately expressed because the word order is not taken into account. To calculate the similarity by focusing on the scenario expressions, the vectorization of documents by word2vec is not appropriate. Then, doc2vec takes into account the context vectors of the scenario. The user's scenario and the pro-scenario are vectorized with doc2vec to calculate cosine similarity.

To research how effective doc2vec is in calculating the similarity of the enterprisers' scenario to the pro-scenario, an

experiment is conducted for 4 subjects. The subjects write scenarios from their own experiences, following the screenwriting technique.

The similarity of the subjects' scenarios to other storylines was calculated. For this purpose, the document vectors of the pro-scenario and the subject's scenario were generated for each storyline. The results of the similarity between the subject's scenario and the pro-scenario for each storyline development are shown in TABLE VIII. The columns on the far left are the scenarios of the four subjects and the top of the most column is the name of the storyline. For example, the similarity between Scenario 1 and "Monster in the House" is 0.138. When the similarity is close to 1.0, the two scenarios compared are similar. The similarity values were lower than 0.2 on the whole. In other words, the subjects' scenarios were not similar to any of the storylines. This is probably because the words used by the subjects and the pro-scenario are not similar. Although doc2vec takes context into account, it is based on word vectors, so the similarity will be low if no common words appear between the two scenarios.

TABLE VIII
COSINE SIMILARITY BETWEEN THE USER'S SCENARIO AND THE PROSCENARIO

	Monster	Fleese	Bottle	Problem	Rites	Buddy	Why	Fool	Institutional	Hero
Text1	0.138	0.143	0.161	0.140	0.132	0.139	0.140	0.138	0.137	0.145
Text2	0.029	0.034	0.034	0.025	0.035	0.033	0.028	0.042	0.041	0.028
Text3	0.051	0.042	0.043	0.047	0.032	0.061	0.042	0.010	0.046	0.053
Text4	-0.125	-0.118	-0.096	-0.124	-0.089	-0.115	-0.107	-0.094	-0.109	-0.134

The scenario is incomplete when the enterprisers begin to write it. Naturally, the enterprisers' unfinished scenario is dissimilar to the pro-scenario. Even if doc2vec succeeds in calculating the similarity, it is impossible to give the enterprisers specific advice on their scenarios. Because of the relationship between the words in the enterprisers' scenario and pro-scenario and the lack of specific advice to the expressions, doc2vec is not appropriate for detecting similarity. Since doc2vec focuses on the document as a whole, it is difficult to find the specific advice in the enterprisers' scenario. The following findings were found in this preliminary experiment. The enterprisers should be given something to rely on to write their scenarios.

To write a scenario, the enterprisers should follow the scene development of the pro-scenario. For improving the enterprisers' scenario to be similar to the pro-scenario, they should notice the words that describe the characteristics of the scene for each scene in the pro-scenario. By dividing the scene into detailed parts, it is expected that the lack of expressions can be found by the existence of words used in the pro-scenario within the enterprisers' scenario.

B. The Similarity Detection with Negative-Positive Analysis

The four scenes of the pro-scenario are written as follows. The first scene is the scene that the story is set up with characters, place, and time. In the second scene, the main character's emotions become temporarily positive due to the events to move the story forward. In the third scene, the main

character's emotions change negatively as the crisis to him. In the fourth scene, the main character's emotions change to positive again because he has overcome the crisis. Although the first scene can be ignored because it is the setting of the story, the second through fourth scenes is written according to the emotional changes of the main character. The main character's emotions change in the following order: positive, negative, and positive. In other words, if the user's scenario is also such a change of emotion, the user can express the scene appropriately. Emotional changes in each scene are not enough to make specific advances about the user's scenario. Therefore, each scene is given positive or negative scores. Positive words are given a positive value. Negative words are given a negative value. This positive and negative score is called the negative-positive score. By comparing the negative-positive scores of each scene in the movie and the user's scenario, it is possible to notice how many expressions are lacked in the user's scenario. The sum of the negative-positive scores for each scene defines whether each scene is positive or negative. For example, a scene with many positive words is defined as a positive scene because the negative-positive score is positive. On the other hand, a scene with many negative words is a negative scene. An emotion polarity dictionary is used to give negative-positive scores to words in a scenario. In the emotion polarity dictionary, many words are given a negative-positive score before using. An emotion polarity dictionary gives negative-positive scores to words in the movie and the user's scenario. For each scene in each scenario, the negative-positive scores of the words are summed up and the average is divided by the number of words in the scene. The negative-positive score of a scene is calculated by the following formula.

Negative - positive score of the scene

$$= \frac{\text{Sum of the negative - positive scores of the words in the scene}}{\text{Number of words in the scene}}$$

The average makes it possible to reduce the effect of differences in the number of words in each scene. If the number of words is large, the negative-positive score may be biased either positively or negatively. The negative-positive scores for each of the ten storyline scenes are shown in Figure 2. The vertical axis is the negative-positive score, and the horizontal axis is the four scenes. The negative-positive score for the first scene of "Monster in the House" is -0.56. The shape of each graph is the change in the negative-positive score for each scene of the story development. The ideal transition of the negative-positive score for the four scene developments is as follows: the negative-positive score decreases from the second scene to the third scene and then increases again toward the fourth scene. The graphs of the expected form are "Golden Fleece" and "Buddy Love," but the graphs of the other storylines are not expected. Because the specific scenes differ in each storyline, the negative-positive scores for each scene are large or small. However, the shape of the graph should be a

V-shape like "Golden Fleece." Since the ten storylines follow the four scene developments of the screenwriting technique, the main character's emotional changes are common in all stories. However, many of the graphs are not expected. There are several reasons for this. The negative-positive score given in the emotion polarity dictionary may not match the meaning of the word used in the actual document. For example, the word "ocean" has a negative score in documents about natural disasters. On the other hand, the documents related to summer vacation and leisure are expected to have positive scores. The emotion polarity dictionary defines the negative-positive score of "sea" as -0.987. As shown above, it is difficult to use an emotion polarity dictionary in a pro-scenario using many words whose positive and negative scores can change with each context. In addition, many undefined words appear in the pro-scenario which are not in the emotion polarity dictionary. These undefined words may have a significant impact on the main character's emotions in the scenario. For this reason, the scenario similarity detection by negative-positive scores using the emotional polarity dictionary is inappropriate.

If the negative-positive score could be used to calculate the similarities between the pro-scenario and the enterprisers' scenario, the following issues would still exist. It is difficult to reevaluate the expressions of the scenario after the improvement by the enterprisers. By comparing the negative-positive scores for each scene in each scenario, the lack of expressions in the user's scenario is specified. For example, in the second scene, the negative-positive score of the pro-scenario is 0.600 and the negative-positive score of the enterprisers' scenario is 0.500. The user writes positive expressions so that his negative-positive score of the second scene becomes 0.600. However, the negative-positive score for each scene is averaged the number of words in the scene. There is little change in the negative-positive score even if the user writes a few more words. This means that it is difficult for users to specify the points of improvement in the scenario. Therefore, the detection of lack of expressions with negative-positive scores is inappropriate. The similarity detection using negative-positive scores had the risk of ignoring undefined words that may have important meanings in the context.

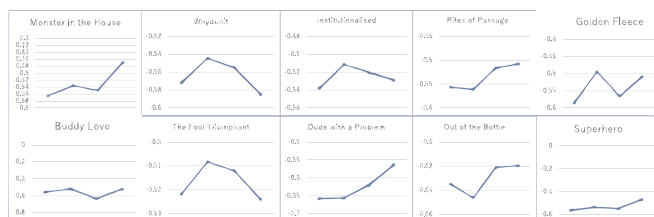


Fig. 2. Negative-positive score for each storyline

The following method can solve this problem. By giving the meaning from the words in the string close to the undefined word, the original meaning of the undefined word may be given in the context. It should be possible to notice the specific lack of expressions in the enterprisers' scenario by paying attention to the words that describe the characteristics of the scene for each scene.

VII. CONCLUSION

In this study, a logistic regression model was created to classify the scenes of the target scenario. The appropriateness of the scene words was tested by creating a concrete model with "Monsters in the House". Based on the coefficients of the scene words in each scene, the selection of appropriate scene words was important for classifying the scenes. In the target scenario, scene words should be used to describe the characteristics of each scene.

Since the content to be written in a scene is different in each story development, the scene words are considered to be different depending on the story development. Therefore, enterprisers should correctly select the storyline that corresponds to the experience. By selecting a storyline that corresponds to the experience and using appropriate scene words, the enterprisers can create a scenario that will gain the reader's empathy.

By testing the appropriateness of the scene words in "Monster in the house," the scene words can be extended to other storylines. However, it is necessary to verify that it is possible to create models with the scene words for all storylines. The next research is to create a model that can select scene words in other storylines and classifies the scenes in the target scenario.

REFERENCES

- [1] T. Y. CHUN, D. K. LEE, and N. H. PARK, "The effect of marketing activities on the brand recognition, brand familiarity, and purchase intention on the sns of franchise companies," *The Journal of Asian Finance, Economics, and Business*, vol. 7, no. 11, pp. 955–966, 2020.
- [2] G. Freytag, *Technique of the drama: An exposition of dramatic composition and art*. S. Griggs, 1895.
- [3] A. G. Woodside, S. Sood, and K. E. Miller, "When consumers and brands talk: Storytelling theory and research in psychology and marketing," *Psychology & Marketing*, vol. 25, no. 2, pp. 97–145, 2008.
- [4] T. Van Laer, S. Feiereisen, and L. M. Visconti, "Storytelling in the digital era: A meta-analysis of relevant moderators of the narrative transportation effect," *Journal of Business Research*, vol. 96, pp. 135–146, 2019.
- [5] S. Blake, *Save the Cat: The Last Book on Screenwriting You'll Ever Need*. Michael Wiese Productions, 2005.