

27th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2023)

Estimating Story-events of Comics Based on Characteristic Words

Taketo Fujikawa^a, Mitsunori Matsushita^a

^a*Kansai University, Ryozenji 2-1-1, Takatsuki 569-1095, Osaka, Japan*

Abstract

This study aims to clarify the composition of story content to help computers understand stories. In story content, the events peculiar to a genre occur intermittently. For example, school and sports festivals appear in school-themed genres. These events can trigger a story because they cause changes in the internal characteristics and relationships between characters, which in turn trigger the progress of the story. If computers can determine the events in a story, they will help understand its composition. Each story-event contains many strongly related words. For example, “relay” and “runner” appear in sports festival episodes. Therefore, investigating these tendencies is expected to contribute to the estimation of story-events. However, the amount of information obtained from comic texts is limited because they use illustrations and texts in a complementary manner. This makes it difficult for computers to obtain words from comics that characterize a story-event. To address this problem, we focused on the content similarities between comics and light novels. In this study, we estimated story-events in comics using the tendency of story-event words to appear in light novels. The results of this experiment suggest that the computer can calculate story-events in comics using a dictionary containing characteristic words of story-events and the proposed method.

© 2023 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the scientific committee of the KES International.

Keywords: story-event; Characteristic words; Estimation of a story-event; Light novel;

1. Introduction

Story content, such as novels, comics, and games, are usually categorized into patterned genres, such as action, fantasy, and mystery. Even if stories are categorized into the same genre, their compositions are not necessarily the same. For example, in the action genre, some works have content in which the main character is the strongest from the beginning of the story and overwhelms consecutive enemies by force. Conversely, other stories have content in which the main character, who is not strong at the beginning of the story, becomes strong by overcoming various difficulties. If a computer can understand and classify the composition of stories, services such as automatically generating stories,

* Corresponding author. Tel.: +81-72-690-2437; fax: +81-72-690-2491.

E-mail address: {k920357, m_mat}@kansai-u.ac.jp

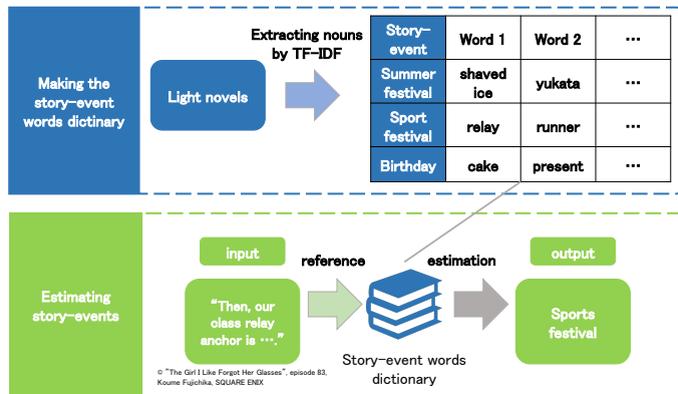


Fig. 1. Overview of this paper

searching for stories by content, and recommending stories that users may prefer, will become possible. In fact, the search system was developed based on the similarity of picture books' story structure [8].

As a step toward helping computers understand the structure of a story, this study attempts to estimate the types of events described in the story. Each genre contains intermittent story-events specific to itself. For example, the action genre includes battles and training events, and the school-themed genre includes school events (e.g., sports festivals and periodic examinations) and annual events (e.g., Christmas and St. Valentine's Day).

These story-events are often used as points of change in the abilities of a character, internal characteristics, and the relationships between characters; such as an episode in which a heroine accidentally sees the hero receiving chocolate from someone other than herself on Valentine's Day and becomes aware of her romantic feelings.

Several studies were conducted on story structure analysis of story content such as comics [2, 5] or novels [4], picture books [8].

2. Research approach

This study focuses on words appearing in story-events as an approach to estimating the type of story-event. In the school-themed genres, "fireworks," "yukata," and "shaved ice," often appear in episodes showcasing a summer festival. Similarly, "relay," "runner," and "hachimaki headband," often appear in sports festival event episodes. This study defines these words characterized by a story-event as "story-event words." Assuming that the text of a story-event tends to have many story-event words, this study gathers them and attempts to make a computer estimate of story-events using them as clues.

A possible primary source for collecting story-event words is the information in the text that appears in the work. However, comics are not suitable for picking story-event words because they use illustrations and text in a complementary manner [6], which complements information that describes a situation and scene about the backgrounds and visual expressions of characters, making comics have less text. Textual information in comics, such as dialogues and monologues, is insufficient for collecting words for story-event estimation. Hence, we focused on light novels, which are young adult novels that primarily target teenagers. Light novels are textual and handle content similar to that in comics. In light novels, in addition to dialogues and monologues, information pictured in comic illustrations is described by texts as descriptive narratives, making it easier to collect vocabulary from light novels than from comics.

Another advantage is that light novels are divided into genres using the same categories as comics, such as battle, fantasy, and romantic comedy. Furthermore, some popular light novels have been adapted as comic books. Therefore, more story-event words should be collected from light novels than from comics. This study attempts to create a word dictionary classified by story-events.

Fig. 1 presents an overview of the study. The proposed method comprises three steps. First, we investigated the story-events in the target genres of light novels and comics. Second, we gathered story-event words from light novels

and created a story-event words dictionary to estimate the story-events. Finally, using the dictionary, we estimated the story-events in light novels and comics.

3. Investigating story-events

This section investigates the number of story-events for each target work and frequency of story-events in comics and light novels to determine the target genre.

3.1. Target genre

The story genre addressed in this study was school-themed for two reasons.

First, the school-themed genre includes more story-events. Second, story-events in school-themed genres can be easily differentiated from other stories. The target school-themed genres concerned the school lives of characters in modern times. Even if a story occurs in a school, works belonging to genres such as action and science fiction are excluded because they contain words and scenes that are not used in daily life.

3.2. Criterion for investigating story-events

First, we chose 21 light novels from “Shousetsuka ni narou” (Let’s Become a Novelist), a novel posting site on the internet, and 19 comics to decide the target story-event arising in the school-themed genre. The following conditions were employed to select 40 works from the school-themed genre.

- The time of stories is in the present day.
- The story occurs in a school, such as a junior high school and high school.
- The timeframe of the story spans more than one year or the story is complete as of October 10, 2022.

We chose works in which the story timeframe spans more than one year because each similar story-event often occurs yearly in school-themed genres and we can exhaustively explore the occurrence of story-events. Story-events, such as school and annual events, are likely to occur multiple times in a single work. For example, in a story describing school life over three years, some story-events may be described twice or thrice. Therefore, we counted the frequency and number of works with event occurrences when counting the number of story-events.

We did not count it as reoccurrences in a number of episodes but as one instance when the same label of the story-event occurred over multiple episodes. Moreover, we counted it one instance when the unrelated episode was between two story-events of the same name, preparing for the school festival, and on the same day. However, even if the story-events had the same name, we counted it as the number of episodes when a story-event was not related to the pre-event of the same name or was a different story-event because of the characters going to the next grade.

3.3. Result of the investigation on story-events

By counting Section 3.2, we found 51 types of story-events that occurred in 40 school-themed genre works. In story-events occurring more than ten times, we found 21 types of events by work and 24 by frequency. Table 1 presents the results of this investigation. In addition, the following are examples of story-events confirmed by more than two works or times.

Ball games, Camping school, Chorus competition, Entrance ceremony, School council election, Studying outdoor, Spring vacation, Winter vacation, Body measurement and sport test, Changing seats in the classroom, Cleaning in school, Cooking practice, Remedial class, Swimming in a school pool, Survey of career choices, April Fool, Festival of the Weaver, Halloween, Mother’s Day, Setsubun: the day before the calendric beginning of spring, Courage test, (parents’) Homecoming visit, Viewing of cherry blossoms, Going to an aquarium, Going to a movie theater, Going to a zoo, Sleepovers

Table 1. Number of story-events

(a) work				(b) frequency			
Name of story-event	Light novel	Comics	Total	Name of story-events	Light novel	Comics	Total
School festival	17	15	32	School festival	20	23	43
Valentine's Day	13	13	26	Birthday	16	22	38
Summer festival	14	10	24	Study group	14	18	32
Christmas	12	12	24	Valentine's Day	16	16	32
Study group	12	9	21	Christmas	15	16	31
Summer vacation	11	10	21	Shopping	12	19	31
Birthday	8	13	21	Summer festival	15	13	28
Sports festival	8	11	19	Summer vacation	13	11	24
School trip	10	8	18	Travel	11	11	22
Shopping	10	8	18	Sea bathing	8	14	22
Sea bathing	6	12	18	Sports festival	8	13	21
New Year's shrine visit	8	7	15	Going into an arcade	8	13	21
White Day	5	10	15	School trip	11	9	20
Periodic examinations	9	5	14	Periodic examinations	12	5	17
Travel	8	6	14	White Day	6	11	17
Swimming in a pool	7	6	13	Swimming in a pool	10	6	16
Graduation ceremony	6	7	13	New Year's shrine visit	8	8	16
New Year's Eve	6	5	11	Graduation ceremony	8	7	15
Going to an amusement park	6	5	11	Nursing for get-well	5	10	15
Going into an arcade	5	6	11	Going to an amusement park	8	6	14
Nursing for get-well	3	8	11	Going to a park	1	11	12
				Going to karaoke	7	5	12
				New Year's Eve	6	5	11
				New Year's Day	4	7	11

Table 2. List of light novels used for making story-event words dictionary

Title ID	Author	URL
LN01	Hayato Inugi	https://ncode.syosetu.com/n1097hj/
LN02	Yuki Ogata	https://ncode.syosetu.com/n5464hl/
LN03	Jiro Kinno	https://ncode.syosetu.com/n2466hh/
LN04	Mumin	https://ncode.syosetu.com/n5377hf/
LN05	Yu Yuki	https://ncode.syosetu.com/n5598hi/
LN06	Kennoji	https://ncode.syosetu.com/n4976ea/
LN07	Moko Haruyo	https://ncode.syosetu.com/n1842hn/
LN08	Mutsukiuduki	https://ncode.syosetu.com/n4926gt/
LN09	Ryo Nishifuji	https://ncode.syosetu.com/n8318hm/
LN10	Sametaro Fukada	https://ncode.syosetu.com/n0674fu/
LN11	[] (Mumei)	https://ncode.syosetu.com/n4284hq/

(Retrieved Feb. 24,2023)

Based on this investigation, we classified story-events occurring in school-themed genres into the following five categories:

1. Actual school events (e.g., school festival, sports festival, school trip)
2. School life events that almost all students experience (e.g., periodic examinations, cooking practice, and changing seats in the classroom)
3. Annual events (e.g., Valentine's Day, Christmas, Halloween)
4. Seasonal events (e.g., summer festival, sea bathing, viewing of cherry blossoms)
5. Events not related to the season outside school (e.g., birthday, study group, going out)

Table 3. Story-events used for making story-event words dictionary and the number of the texts

Name of story-event	The number of text
School festival	6,019
School trip	3,041
Summer festival	2,260
Sports festival	2,108
Christmas	1,804
Swimming in a pool	1,395
Travel	1,291
Birthday	1,281
Shopping	1,215
Study group	988
Nursing for get-well	934
Ball game cup	669
Graduation ceremony	556
Valentine's Day	376
Going into an arcade	355
New Year's Eve	334
New Year's shrine visit	333
Going to karaoke	254
Going to a theater	220
Periodic examination	144
permit duplication	

4. Story-event words dictionary

We created a story-event words dictionary to decide in which story-events the story-event word needed to estimate story-events appears.

4.1. Preprocessing

We chose 11 light novels for making a story-event words dictionary from “Shousetsuka ni narou” (Let’s Become a Novelist), as summarized in Table 2. In addition to the conditions described in Section 3.2, we chose light novels, each with an episode title because an episode title makes it easy to identify the story-events that occur in the episode.

We labeled the entire text of the title as a story-event (31,240 sentences) and in the gathered texts (92,937 sentences), the name of the story-event. The study included text (28,618 sentences) labeled with the terms of the top 20 story-events that occurred in the 11 works, excluding “summer vacation,” which occurred more than three times. Table 3 summarizes the number of texts labeled with the names of each story-event. Note that we excluded “summer vacation” because it included many story-events such as “summer festival,” “swimming in a pool,” and “sea bathing.”

We conducted a morphological analysis of texts corresponding to the 20 story-events to extract nouns. We excluded other independent words, such as verbs and adjectives, from this study because they are less likely to appear than nouns depending on a specific story-event. There was a concern that they may affect the accuracy of story-event estimation.

We used Mecab (version 0.996)¹ [3] as a morphological analyzer and ipadic-Neologd² as a dictionary for morphological analysis. The words included in Slothlib³, single-character hiragana and katakana, symbols, names of people, and personal pronouns unrelated to story-event estimation, were adopted as stop words. Morphological analysis using the ipadic-Neologd dictionary yielded 10,460 nouns.

¹ <https://taku910.github.io/mecab/> (Retrieved Feb. 24,2023)

² <https://github.com/neologd/mecab-ipadic-neologd/blob/master/README.ja.md> (Retrieved Feb. 24,2023)

³ <http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word/Japanese.txt> (Retrieved Feb. 24,2023)

Table 4. Part of the result of extracting story-event words
(a)summer festival

Words	The value of TF-IDF
wheelchair	0.413
boy	0.344
firework	0.332
yukata	0.219
summer festival	0.205
red	0.196
Hyottoko	0.189
festival	0.174
food stall	0.165
Senior	0.121

(b)school trip

Words	The value of TF-IDF
girl	0.316
chairperson	0.305
school trip	0.254
teacher	0.173
Dad	0.155
senior	0.137
room	0.130
Two	0.129
Mom	0.124
like	0.122

Table 5. Number of story-event words added to each dictionary in target story-event

	School festival	Valentine's Day	Summer festival	Christmas	Study Group	Birthday	Sports festival	School trip	Shopping	New Year's shrine visit
Not refined dictionary	50	50	50	50	50	50	50	50	50	50
Refined dictionary	24	12	22	10	21	9	31	14	13	27

4.2. Extracting story-event words with TF-IDF

We attempted to extract story-event words by TF-IDF because these words frequently only appear in some story-events. Table 3 presents the TF-IDF values for the 20 story-events for the nouns extracted in Section 4.1. The TF-IDF values were calculated using the TfidfVectorizer in the sklearn.feature_extraction.text class of Scikit-learn (version 1.1.1)⁴. The results are presented in Table 4.

Table 4 (a) shows that many story-event words, such as “fireworks,” “yukata,” and “summer festival,” appear among the top 10 words with the highest TF-IDF value in summer festivals. However, Table 4 (b) lists story-event words such as “souvenir,” “sightseeing,” and “group,” which are closely related to school trips and did not appear often.

4.3. Making story-event words dictionary

In this study, we created two types of story-event word dictionaries: a non-refined dictionary and a refined dictionary. The non-refined dictionary contained the top 50 words with the highest TF-IDF values calculated for each story-event in Section 4.2 treated as story-event words. In the refined dictionary, we manually selected story-event words considered relevant to the story-event from the top 50 words. The following are examples of words included in the refined dictionary for summer festivals.

fireworks, yukata, summer festival, Hyottoko (Japanese traditional mask), festival, food stall, fried noodles, park, site, geta (Japanese traditional clogs), stage, shaved ice, babel, goldfish, Japanese drum, takoyaki, goldfish scooping, cotton candy, river location, happi coat

Table 5 summarizes the number of words used for each story-event in the non-refined and refined dictionaries.

5. Proposed method for the story-event estimation

Generally, light novels and comics are separated by episodes. It is rare for more than two story-events to occur in one episode; therefore, we made the computer estimate story-events by single episodes. The story-events targeted for estimation in this study were the 20 story-events in Table 1 used to create the story-event words dictionary and the

⁴ <https://scikit-learn.org/stable/> (Retrieved Feb. 24,2023)

Table 6. Parameter selection of the full-text (a) with the non-refined dictionary

p	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$
0.30	0.443	0.420	0.404	0.386	0.371	0.367
0.40	0.251	0.268	0.292	0.315	0.331	0.347
0.50	0.115	0.143	0.207	0.221	0.259	0.267
0.60	0.032	0.059	0.097	0.139	0.195	0.241

(b) with the refined dictionary

p	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$
0.30	0.336	0.395	0.448	0.506	0.547	0.555
0.40	0.363	0.402	0.447	0.505	0.549	0.557
0.50	0.355	0.396	0.439	0.504	0.548	0.555
0.60	0.379	0.411	0.453	0.502	0.541	0.546

top 10 story-events among them: “school festival,” “Valentine’s day,” “summer festival,” “Christmas,” “study group,” “birthday,” “sports festival,” “school trip,” “shopping” and “New Year’s shrine visit.”

For each of the 11 works used to create the story-event words dictionary in Table 2, we utilized MeCab to split the text contained in each episode into morphemes and extract nouns.

The computer identified the story-events to which the extracted nouns belonged by referring to a story-event words dictionary. Then, for each story-event, the computer counted the number of appearances of the nouns identified as a story-event, allowing for duplication. Based on the results of the counting, the computer created a 10-dimensional vector $E_a(i)$ (hereafter referred to as “story-event vector”) of 10 different story-events (e_1, e_2, \dots, e_{10}) for episode i ($i = 1$ to T) of work a ($a = 0$ to W).

$$E_a(i) = \{e_1, e_2, \dots, e_{10}\}. \quad (1)$$

We set the cutoff criterion N because of the nature of the method for estimating story-events from one episode; story-event words that are not related to the original story-event may accidentally appear. For example, other story-event words, such as “return of *Valentine’s Day*” appear in one episode of “*White Day*.” This is why other story-event words may become noisy in story-event estimation using this method. The computer assigned zeros to the elements of the story-event vector in which the number of appearances of the story-event word was less than the cutoff criterion. The story-event vectors obtained were normalized such that the sum of their elements was 1.0.

$$e'_j = \frac{e_j}{\sum_{k=1}^{10} e_k}. \quad (2)$$

$$E_a(i) = \{e'_1, e'_2, \dots, e'_{10}\}. \quad (3)$$

If the element of the story-event vector with the highest value exceeded the threshold value, it was adopted as the story-event estimation.

6. Experiments

This section describes the pre-experiment conducted to set two parameters: the cutoff criterion N and story-event estimation threshold p , described in Section 5, and the story-event estimation results for each of the light novels and comics conducted using the set parameters.

6.1. Experiment 1: parameter selection for cutoff criterion and event estimation threshold

We experimented to set the cutoff criteria and story-event estimation thresholds. The experiment was conducted in three steps. First, the computer estimated a story-event in each episode of the 11 light novels (Table 2), which was used to create the story-event words dictionary. The computer used all the text in one episode (hereafter referred to as full-text) for story-event estimation. Second, the computer calculated the recall, precision, and F-score, to evaluate the story-event estimation accuracy. Finally, we selected the combination of N and p with the highest calculated F-score based on a grid search. N was set to one increment from one to six, and p was set to 0.10 increments from 0.30 to 0.60.

Table 6 summarizes the results of the non-refined and refined dictionaries constructed in Section 4. The maximum F-score was obtained for the combination of $N = 1$ and $p = 0.30$ for the non-refined dictionary, and $N = 6$ and $p = 0.40$ for the refined dictionary. Story-event estimation with the full-text light novels using the refined dictionary outperformed the refined dictionary by 0.114 points.

Table 7. Parameter selection of dialogue texts
(a) with the non-refined dictionary

p	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$
0.30	0.290	0.286	0.311	0.325	0.299	0.311
0.40	0.242	0.261	0.295	0.317	0.306	0.322
0.50	0.196	0.233	0.273	0.306	0.275	0.305
0.60	0.166	0.187	0.250	0.324	0.297	0.308

(b) with the refined dictionary

p	$N = 1$	$N = 2$	$N = 3$	$N = 4$	$N = 5$	$N = 6$
0.30	0.341	0.383	0.415	0.396	0.353	0.327
0.40	0.349	0.392	0.413	0.393	0.353	0.327
0.50	0.347	0.394	0.414	0.393	0.353	0.327
0.60	0.362	0.381	0.393	0.387	0.339	0.328

Table 8. List of light novels used for this experiment

Title ID	Author	URL
LN12	Yu Taira	https://ncode.syosetu.com/n3483hs/ **
LN13	Sakuya Kuroyuri	https://ncode.syosetu.com/n9803ho/ *
LN14	—	https://ncode.syosetu.com/n6500hi/ *
LN15	Seisui Edo	https://ncode.syosetu.com/n8985hf/ **
LN16	Yoko Kozuki	https://ncode.syosetu.com/n3682fj/ **
LN17	Kyusaku Amenohi	https://ncode.syosetu.com/n6893en/ **
LN18	Peipei	https://ncode.syosetu.com/n0223hd/ **
LN19	Gurepuhiyasinsu	https://ncode.syosetu.com/n0237fj/ **
LN20	Ripuro	https://ncode.syosetu.com/n3156gx/ **
LN21	Sanshokuraito	https://ncode.syosetu.com/n5685ho/ **

*(Retrieved Oct. 1,2022) **(Retrieved Feb. 24,2023)

Additionally, considering that the number of words used as cues for story-event estimation in comics was smaller than that in light novels, we conducted a similar experiment using only dialogue texts from the same 11 light novels. The results are summarized in Table 7. The maximum F-score was obtained for the combination of $N = 4$ and $p = 0.30$ for the non-refined dictionary, and $N = 3$ and $p = 0.30$ for the refined dictionary. Story-event estimation with dialogue texts of light novels using the refined dictionary also outperformed the non-refined dictionary by 0.090 points.

It was confirmed that the accuracy of the story-event estimation of the refined dictionary was higher than that of the non-refined dictionary for the full and dialogue texts of light novels only. This was presumably because the non-refined dictionary contained words considered less relevant to the story-event; they are often used daily. For example, Table 4 (b) in Section 4.2 indicates that the non-refined dictionary for the event “school trip” included words such as “girl,” “chairperson,” and “teacher.” These daily words appear frequently in the text and are easy to extract as story-event words. This results in an increase in the number of story-event words in each story-event per episode. Therefore, words used daily become noise in story-event estimation. If the amount of noise in the story-event words is large, the cutoff criterion will not work even if the value of N is increased. Consequently, multiple elements of the story-event vector can have value, and it is easy to incorrectly estimate whether a story-event that has not actually occurred has occurred in an episode or to estimate a different story-event. This indicates that the accuracy of the story-event estimation using the non-refined dictionary is low.

When the computer estimated story-events, we decided to use the refined dictionary as the story-event words dictionary in the main experiment and adopted a combination of parameters: $N = 6$ and $p = 0.40$ for the light novels and $N = 3$ and $p = 0.30$ for the comics.

6.2. Experiment 2: story-event estimation

Using N and p selected in Section 6.1, we conducted an evaluation experiment on the story-event words dictionary constructed in this study on the texts of ten other light novels and two comics.

6.2.1. Result of story-event Estimation in Light novels

To verify the feasibility of story-event estimation using the story-event words dictionary constructed in this study, we conducted experiments using light novels that were not included in the construction of the story-event words dictionary. Table 8 lists the 10 light novels used in the main experiment.

Story-event estimation in the ten light novels resulted in a recall of 0.465, precision of 0.355, and F-score of 0.403. This result suggests that the proposed method can provide a certain degree of story-event estimation for novels.

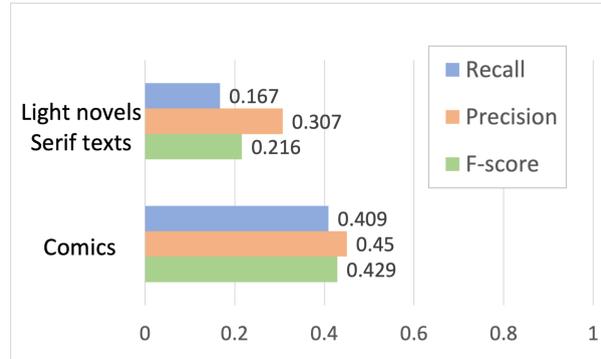


Fig. 2. Comparison results of story-event estimations

Table 9. Result of estimation of appeared story-events in two comics

	School festival	Valentine's Day	Summer festival	Christmas	Study Group	Birthday	Sports festival	School trip	Shopping	New Year's shrine visit
Recall	0.400	1.000	0.250	0.000	—	0.000	1.000	—	—	—
Precision	0.800	0.667	0.500	—	0.000	—	0.500	0.000	—	0.000
F-score	0.533	0.800	0.333	—	—	—	0.667	—	—	—

6.2.2. Result of story-event Estimation in Comics

In this experiment, we targeted two comics: “*The Girl I Like Forgot Her Glasses*”[1] and “*Daily Lives of High School Boys*”[7]. An experiment was conducted to evaluate the accuracy of story-event estimation for $N = 2$ and $p = 0.60$, as described in Section 6.1. Moreover, we compared the story-event estimation results between the two comics and only the dialogue texts of the ten light novels used in Section 6.2.1. Figure 2 presents a comparison of the results.

Story-event estimation for the two comics resulted in a recall of 0.409, precision of 0.450, and F-score of 0.429. This result suggests that the proposed method can provide a certain degree of story-event estimation for comics. Table 9 presents the estimation results for each story-event in the two comics.

6.2.3. Discussion

Figure 2 shows that the recall, precision, and F-score, of the comics were all higher than those of the dialogue texts of light novels. This is presumably because comics often have dialogues called expository dialogues to explain situations and they contain more story-event words than the dialogue texts in light novels.

Table 9 indicates that story-event estimation accuracy was high for two events, Valentine’s Day and sports festivals. When the story-events were correctly estimated, it was confirmed that story-event words that belonged to the story-event appeared more frequently in the episodes. Contrarily, two story-events, Christmas (two episodes) and birthdays (two episodes), were not estimated. This was because of the low story-event words for Christmas and birthdays corresponding stories. The computer incorrectly estimated the story-events of study group (4 episodes), school trip (1 episode), and New Year’s shrine visit (1 episode), not occurring in the two comics. It was caused by story-event words that are also used daily such as “textbook” belonging to study groups, “room” belonging to school trips, and “onegai” belonging to New Year’s shrine visit. “Textbook” is a word that is used daily by students (e.g., in class), leading to incorrect estimation. The word “room” was selected as a story-event word because it is related to school trips (e.g., hotel rooms and room assignments). However, it is also used daily, leading to incorrect estimation. The word “onegai” means “wish” in English. It was selected as a story-event word because it often appears in New Year’s shrine visit. However, “onegai” is used in Japanese when you ask for a favor. Therefore, the computer incorrectly estimates episodes that have the scene asking for a favor.

Based on these, we can improve the accuracy of story-event estimation by using co-occurrence relationships with other story-event words. Among story-event words, words that are easy to use daily set their weight lower. Then, if other story-event words occur with them in the same episode, their weight up. By doing that, the estimated value will be lower in the case of only them in an episode. That makes incorrect estimations by them decrease. On the other hand, we can use both texts and illustrations for story-event estimations of comics. Illustrations have information

about the characters' clothes, the stage of the episode, items, and backgrounds. In the case of a sea bathing episode, the illustration often describes that the characters' clothes are bathing costumes and the stage is a beach. In the case of a summer festival episode, for instance, the illustration often describes yukata as the characters' clothes, fried noodles, and cotton candy as food items, and food stalls and fireworks as the backgrounds. If computers recognize such visually-represented information, they will help with story-event estimation.

7. Conclusions

This study proposes a method for estimating story-events in comics. The results suggest that we can estimate story-events using a dictionary that includes characterized words related to story-events. Their words were gathered from light novels: (1) We clarified story-events according to the categorization of events in the school-themed genre. (2) The results of the experiment for setting the parameters showed that including words unrelated to a story-event in the story-event word dictionary impairs the F-score. (3) The results of the experiment for story-event estimation revealed that estimating story-events in comics is possible using the proposed method. We will increase the number of words included in the story-event word dictionary and use the co-occurrence relationship among story-event words to improve the accuracy of estimating story-events.

Acknowledgement

We would like to thank the authors of the light novels and comics used in this study. This study was supported by a Grant-in-Aid for Scientific Research (22K12338).

References

- [1] Fujichika, K., 2019. *The Girl I Like Forgot Her Glasses*. SQUARE ENIX.
- [2] Imaizumi, K., Yamanishi, R., Nishihara, Y., Ozawa, T., 2021. Estimating groups of featured characters in comics with sequence of characters' appearance, in: International Joint Workshop on Multimedia Artworks Analysis and Attractiveness Computing in Multimedia, pp. 13–18.
- [3] Kudo, T., Yamamoto, K., Matsumoto, Y., 2004. Applying conditional random fields to japanese morphological analysis, in: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Proceeding, pp. 230–237.
- [4] Murai, H., 2017. Automatic extraction of reversal-type punch lines in shin'ichi hoshi's flash fictions, in: Journal of the Japanese Association for Digital Humanities, pp. 31–47.
- [5] Murai, H., 2020. Factors of the detective story and the extraction of plot patterns based on japanese detective comics, in: Journal of the Japanese Association for Digital Humanities, pp. 4–21.
- [6] Yamanishi, R., Matsushita, M., 2017. Comic computing: a conceptual framework for decomposition and utilization of comic contents, in: The 6th Asian Conference on Information Systems, pp. 275–285.
- [7] Yamauchi, Y., 2010. *Daily lives of high school boys*. SQUARE ENIX.
- [8] Yasuo, M., Matsushita, M., Hattori, T., Fujita, S., 2017. Measuring similarity of story development for picture book search, in: The 2017 Conferece on Technologies and Applications of Artificial Inntelligence.