

イベント特徴語に基づくコミックのイベント推定

藤川 雄翔^{†,a} 松下 光範^{†,b}

[†] 関西大学総合情報学部

a) k920357@kansai-u.ac.jp b) m_mat@kansai-u.ac.jp

概要 コミック作品のストーリーはアクションやファンタジーなどのジャンルに分類され、ジャンル毎に特有の共通したイベント（e.g., 学園ジャンルでは体育祭、文化祭）が発生する。そのイベントの記述においては、そのイベントを特徴づける語彙（e.g., 体育祭であれば“リレー”や“騎馬戦”）が集中して出現するケースが多く見られる。そのため、これらの出現傾向を調べることはイベント推定に貢献すると考えられる。しかし、コミック中のテキスト情報はセリフやモノログなど限定的であるため、イベントを特徴づける語彙を効率的に収集することが難しい。そこで本研究ではコミックとライトノベルの内容的類似性に着目し、ライトノベルから収集したイベント特徴語を用いることで、コミックのイベント推定を試みる。実験の結果から、ライトノベルのテキストを用いることにより、コミックのイベント推定に利用可能な語彙辞書を構築可能であることが確認された。

キーワード ストーリーイベント、イベント特徴語、イベント推定、ライトノベル

1 はじめに

小説、コミック、ゲームなど様々な物語コンテンツには、アクションやファンタジーなど典型的なジャンルが存在する。しかし、同一のジャンルに分類される作品同士であっても、そのストーリーの構成は必ずしも共通ではない。例えば、バトルジャンルでは、身体能力や特殊能力に恵まれた最強の主人公が相対した敵を圧倒していくストーリーもあれば、身体能力や特殊能力に恵まれない主人公が修行を積むことで成長し強くなっていく過程を描いたストーリーもある。このようなストーリーの構成を機械的に把握し分類することが可能になれば、物語コンテンツの自動生成や内容に基づく類似検索、ユーザーの嗜好に合ったストーリー構成を持つ物語コンテンツの推薦などといった活用が期待できる。

物語コンテンツのジャンルには、各々のジャンルに特有のストーリーイベント（以下、イベント）が断続的に発生する。例えば、アクションジャンルのストーリーでは、戦闘や修行などがイベントに該当し、学園ジャンルのストーリーでは、文化祭や体育祭のような学校行事やバレンタインやクリスマスのような年中行事などがイベントとして該当する。

こうしたイベントはストーリーの展開や変化を誘引する起点として描写されることが多い。例えば、バレンタインに他の女子からチョコレートをもらって告白をされている主人公をヒロインが偶然目にし、主人公に対する気持ちを自覚するといったポジティブな心情の変化や、クリスマスに突然のアクシデントに巻き込まれ、遊ぶ約束を守ることができず喧嘩をしてすれ違ふといったネガ

ティブな関係性の変化などがある。このように、イベントはキャラクターの心情やキャラクター同士の関係性に変化を生じさせ、ストーリーを駆動するきっかけとなるため、イベント推定が可能となれば、ストーリーの構成把握に繋がるのが期待される。

ストーリーの構成を把握するために、ストーリーの構造を抽出・分析する研究が様々な媒体を対象に行われてきている [1, 2, 3]。これらの物語構造分析に関する研究はストーリーを人手で分節化し機能に分けることでストーリーを表現しているが、これを人手で行うことは負担になる。そこで、本研究では、ストーリーを駆動するきっかけとなるイベントを計算機により推定し分節化することでストーリーの構成把握の容易化を目指す。その端緒として本稿では、イベントに出現した単語を手がかりにしたイベント推定を試みる。

2 本研究のアプローチ

本研究はイベントを推定するためのアプローチとして、イベントに出現する単語に着目した。学園ジャンルで発生するイベントである夏祭りでは、“花火”、“浴衣”、“かき氷”など、体育祭では、“リレー”、“借り物競走”、“騎馬戦”などといった単語が出現する。本研究では、このようなイベントに関係のある単語を“イベント特徴語”と定義する。発生したイベントのテキストにはそのイベントに関係のある単語が多く出現するという仮定に基づき、イベント特徴語を収集し、それを手がかりとすることで機械によるイベント推定を試みる。

イベント特徴語の収集においては作品中に出現するテキスト情報がその収集源として期待される。コミックは、絵と文字が相補的かつ協調的に利用されているクロ

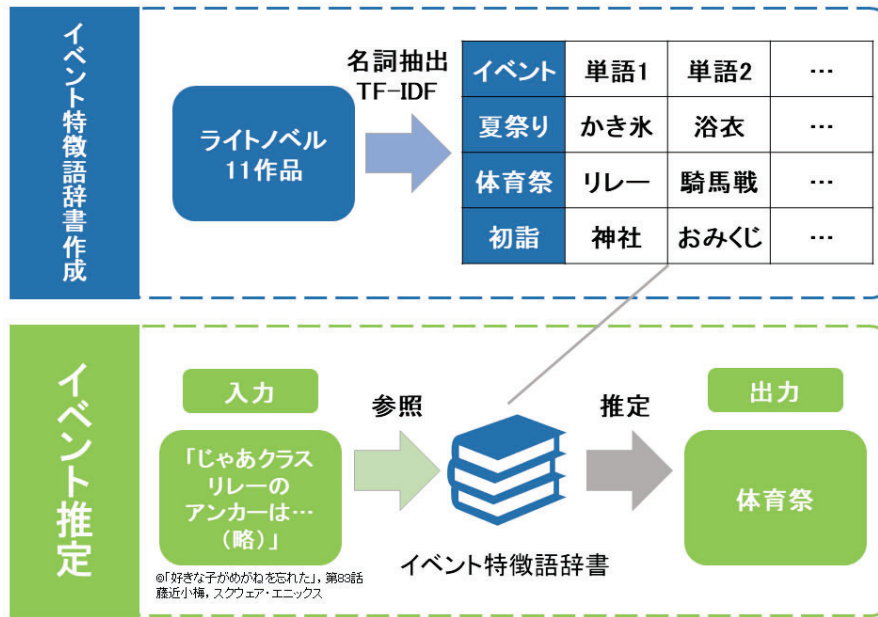


図1 本稿の概要

スモーダルなコンテンツであり [4], 絵として描かれる背景情報や表情が場面や状況の情報を補完するため, セリフやモノログなどのテキスト情報だけではイベントに関する情報の取得に大量のコンテンツが必要になると懸念される. 一方, 小説の場合は, コミックにおいてイラストで描写されている情報が地の文として表現されているため, 小説はコミックよりもテキスト情報の取得が容易である. そこで本稿では, コミックの類似コンテンツのひとつとして小説の一種であるライトノベルに着目した. ライトノベルはコミックと同じようにバトルやファンタジー, 恋愛などといったジャンルに分類され, 人気のあるライトノベルはコミカライズされることが多い. 小説の方がコミックよりも多くのイベント特徴語が抽出できると推定されるため, 本研究ではライトノベルの小説文からイベント特徴語辞書の作成を試みる.

本研究の概要を図1に示す. 本稿ではまず対象となるコミックやライトノベルで発生するイベントについて調査する. 次に, イベント推定のためライトノベルからイベント特徴語を抽出し, イベント特徴語辞書を作成する. 最後に, 作成したイベント特徴語辞書を用いて, ライトノベルとコミックを対象に提案手法によるイベント推定を行う.

3 イベントの調査

本章では, 本稿で対象とするストーリージャンルを決定する. また, 対象とするイベントを定義するため, 決定したジャンルに属するコミックやライトノベルにおいて, 同一名のイベントが発生した作品数とイベントの発

生頻度を人手により調査を行う.

3.1 対象とするストーリージャンル

本稿で対象とするストーリージャンルは他のジャンルよりもイベントの種類が豊富でイベントの判別が容易だと考えられる学園ジャンルとする. 対象とする学園ジャンルは, 時代設定が現代かつメインのキャラクター達の学校生活を中心として描いている題材の作品とする. これは, 学園が舞台となっているアクションやSFなどのジャンルの作品を研究対象として扱わないことで, 非日常的な単語やシーンが出現することを防ぐためである.

3.2 イベントの調査方法

まず, 本研究で対象とするイベントを決めるために, 学園ジャンルの作品を Web 小説投稿サイト「小説家になろう」¹から 21 作品, コミックから 19 作品選定し, 2 作品以上で発生したイベントについて調査した. 調査対象とした 40 作品の学園ジャンル作品は

- 時代設定が現代である
- 舞台が学校・学園である
- ストーリー上の時間軸が1年以上経過している, もしくは 2022 年 10 月 10 日時点で既に完結済みである

という条件を満たす作品とした. ストーリー上の時間軸が1年以上経過している作品を選定したのは, 1年周期で類似したイベントが発生し, イベントを網羅的に収

¹<https://syosetu.com/> (2023/2/24 存在確認)

表1 対象作品中で発生したイベント数 (作品毎)

イベント名	小説	コミック	合計
文化祭	17	15	32
バレンタイン	13	13	26
夏祭り	14	10	24
クリスマス	12	12	24
勉強会	12	9	21
夏休み	11	10	21
誕生日	8	13	21
体育祭	8	11	19
修学旅行	10	8	18
ショッピング	10	8	18
海水浴	6	12	18
初詣	8	7	15
ホワイトデー	5	10	15
定期考査	9	5	14
旅行	8	6	14
プール	7	6	13
卒業式	6	7	13
大晦日	6	5	11
遊園地	6	5	11
ゲームセンター	5	6	11
お見舞い・看病	3	8	11

表2 対象作品中で発生したイベント数 (発生頻度)

イベント名	小説	コミック	合計
文化祭	20	23	43
誕生日	16	22	38
勉強会	14	18	32
バレンタイン	16	16	32
クリスマス	15	16	31
ショッピング	12	19	31
夏祭り	15	13	28
夏休み	13	11	24
旅行	11	11	22
海水浴	8	14	22
体育祭	8	13	21
ゲームセンター	8	13	21
修学旅行	11	9	20
定期考査	12	5	17
ホワイトデー	6	11	17
プール	10	6	16
初詣	8	8	16
卒業式	8	7	15
お見舞い・看病	5	10	15
遊園地	8	6	14
公園	1	11	12
カラオケ	7	5	12
大晦日	6	5	11
正月	4	7	11

集することができると考えたためである。また、ストーリー上の時間軸が1年以上経過している作品を対象としたため、1年毎に発生する学校行事や年中行事のようなイベントが1作品の中で複数回発生する可能性があるため、イベントが発生した作品数も調査することとした。

複数話にわたって同一名のイベントが発生していた場合は、話数分を数えるのではなく、1回とした。また、文化祭の準備がメインの話と文化祭本編の話の間に文化祭とは関係のない話が含まれているような場合は、文化祭の準備と本編、2つの話を同一のイベントとみなし1回のみとしている。ただし、同一名のイベントであっても、前回の同一名イベントとつながりがない、もしくは学年が異なるといった場合は発生した回数を計数した。

3.3 調査によって得られたイベント

3.2節で述べた数え方により、学園ジャンル40作品で発生したイベントは51種類確認された。そのうち、10回以上発生したイベントは作品単位での計数では21種類、発生頻度では24種類が確認された。分布と頻度についてのイベントの調査結果を表1と表2に示す。

今回の調査をもとにイベントを精査し、本稿で対象とする学園ジャンルのイベントを以下の5種にあてはまるイベントに定めた。

1. 現実にも存在する学校行事イベント (e.g., 文化祭, 体育祭, 修学旅行)
2. 学校行事ほどの規模ではないが学園生活を送っていればほとんどの学生が経験する学校生活イベント (e.g., 席替え, 定期考査, 調理実習)
3. 季節に関わる年中行事イベント (e.g., バレンタイン, ハロウィン, クリスマス)
4. 四季に依存して発生する季節イベント (e.g., 花見, 夏祭り, 海水浴)
5. 学園ジャンルのコミックで複数の作品において確認されたキャラクターの個人的イベント (e.g., おでかけ, 誕生日, 勉強会)

4 イベント特徴語辞書

イベント推定に必要なイベント特徴語がどのイベントに属するのか定義するためにイベント特徴語辞書を作成する。

表3 イベント特徴語辞書作成に用いた小説一覧

作品名 (URL)	作者名
貴女が死ぬまでは傍にいる (https://ncode.syosetu.com/n1097hj/)	犬議隼人
ウツつきたちの恋—幼馴染みが俺のことを忘れたみたいなので、幼馴染みの友達と仲良くなってみた— (https://ncode.syosetu.com/n5464hl/)	おがた悠希
オタクな彼女は好きですか? (https://ncode.syosetu.com/n2466hh/)	金野次郎
織田信長が女子に生まれ変わったけど、どうやら穏やかに生きられない! (https://ncode.syosetu.com/n5377hf/)	夢生明
金髪碧眼の女子バレー部のエースに告白したらフラれたけど、俺が作った弁当だけは毎日食べたいと懇願された話。 (https://ncode.syosetu.com/n5598hi/)	悠木ゆう
高2にタイムリープした俺が、当時好きだった先生に告った結果 (https://ncode.syosetu.com/n4976ea/)	ケンノジ
サボり魔先輩に恋に落ちるまで (https://ncode.syosetu.com/n1842hn/)	春夜もこ
スノーフレークと君影草〜不器用な雪片くんとさみしがり屋の鈴蘭ちゃん〜 (https://ncode.syosetu.com/n4926gt/)	睦月卯月
伝えられない想いは空を舞う (https://ncode.syosetu.com/n8318hm/)	西藤りょう
やたらと祭りのいい俺は、毒舌クーデレ美少女の小さなデレも見逃さずにぐいぐいいく (https://ncode.syosetu.com/n0674fu/)	ふか田さめたろう
連勝中の私が唯一勝てないもの (https://ncode.syosetu.com/n4284hq/)	「」(むめい)

(2023/2/24 存在確認)

4.1 前処理

イベント特徴語辞書を作成するにあたって、Web 小説投稿サイト「小説家になろう」から 11 作品を選定した。選定した 11 作品を表 3 に示す。

選定基準は 3.2 節で述べた条件のうち、さらにイベントの特定が容易なタイトル (e.g., 「凧と人気者と妹、プールに行く。」² 「後 1 年〜最悪の卒業式〜」³, 「桜聖祭とグループ決め」⁴) が割り振られている作品とした。収集したテキスト (92,973 文) のうち、イベントを特定することのできたタイトル全てのテキスト (31,240 文) にイベント名のラベル付けを行った。今回、11 作品に発生したイベントのうち、発生頻度が 3 回以上であった“夏休み”を除く上位 20 種類のイベント名がラベル付けされた文章 (28,618 文) を対象とした。表 4 にイベントにラベル付けされた文章数を示す。なお、“夏休み”イベントにはイベントの中に“夏祭り”や“プール”、“海水浴”といった多くのイベントが内包されるため、除外することとした。

その後、20 種のイベントに該当したテキストに対して形態素解析を行い、名詞を抽出した。動詞や形容詞などの他の自立語は名詞に比べて特定のイベントに依存して出現することが少なく、イベント推定の精度に影響を与えることが懸念されるため本研究では除外した。

形態素解析には Mecab (ver. 0.996)⁵を採用し、辞書

には ipadic-NEologd⁶を用いた。ストップワードは Slothlib⁷に含まれている単語と 1 文字のひらがななどカタカナ、記号に加えて、イベント推定に関係のない人名や人称代名詞とした。ipadic-NEologd 辞書を用いた形態素解析により、抽出した名詞の異なり数は 10,460 語であった。

4.2 TF-IDF によるイベント特徴語の抽出

イベント特徴語はそのイベント内で頻出し、かつ他のイベントにはあまり出現しないような単語となっていることから、TF-IDF によるイベント特徴語の抽出を試みた。4.1 節で抽出したそれぞれの品詞の単語について、表 4 の 20 種類のイベントで TF-IDF 値を求めた。TF-IDF 値の算出には Scikit-learn (ver. 1.1.1)⁸のクラスである `sklearn.feature_extraction.text` の `TfidfVectorizer` を使用した。結果の一部を表 5 に示す。

表 5(a) より、夏祭りにおいて、名詞では“花火”、“浴衣”、“夏祭り”などといった夏祭りというイベントを表すようなイベント特徴語が TF-IDF 値の上位 10 単語の中だけでも多く現れていることが確認できる。一方、表 5(b) の修学旅行では、修学旅行というイベントと関係の深い“お土産”、“観光”、“グループ”などといったイベント特徴語はあまり出現しなかった。

⁶<https://github.com/neologd/mecab-ipadic-neologd/blob/master/README.ja.md> (2023/2/24 存在確認)

⁷<http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word/Japanese.txt> (2023/2/24 存在確認)

⁸<https://scikit-learn.org/stable/> (2023/2/24 存在確認)

²オタクな彼女は好きですか?, 第 42 話

³サボり魔先輩に恋に落ちるまで, 第 1 話

⁴伝えられない想いは空を舞う, 第 13 話

⁵<https://taku910.github.io/mecab/> (2023/2/24 存在確認)

表4 イベント特徴語辞書の作成に用いたイベントとその文章量

イベント名	文章量
文化祭	6,019
修学旅行	3,041
夏祭り	2,260
体育祭	2,108
クリスマスイブ・クリスマス	1,804
プール	1,395
旅行	1,291
誕生日	1,281
ショッピング	1,215
勉強会	988
お見舞い・看病	934
球技大会	669
卒業式	556
バレンタイン	376
ゲームセンター	355
大晦日	334
初詣	333
カラオケ	254
映画館	220
定期考査	144
重複あり	

4.3 イベント特徴語辞書の作成

今回、4.2節でそれぞれのイベントにおいて算出したTF-IDF値の高い上位50単語をイベント特徴語とした非洗練辞書と、上位50単語の中から人手でそのイベントに関連すると考えられるイベント特徴語の選定を行った洗練辞書の2種類のイベント特徴語辞書を作成した。以下に夏祭りイベントの洗練辞書に含まれる語彙の例を示す。

“花火”, “浴衣”, “夏祭り”, “ひよっこ”, “祭り”, “屋台”, “お祭り”, “焼きそば”, “公園”, “会場”, “下駄”, “ステージ”, “かき氷”, “喧騒”, “金魚”, “太鼓”, “たこ焼き”, “露店”, “金魚すくい”, “わたあめ”, “河川敷”, “法被”

また、表6に、それぞれのイベントにおけるイベント特徴語辞書の単語数を示す。

5 イベント特徴語辞書を用いたイベント推定手法

一般にライトノベルやコミックには話の単位で区切りが存在する。複数のイベントが1話に発生することは少

表5 イベント特徴語の抽出 (一部)

単語	TF-IDF 値	単語	TF-IDF 値
車椅子	0.413	少女	0.316
少年	0.344	委員長	0.305
花火	0.332	修学旅行	0.254
浴衣	0.219	先生	0.173
夏祭り	0.205	パパ	0.155
レッド	0.196	先輩	0.137
ひよっこ	0.189	部屋	0.130
祭り	0.174	ふたり	0.129
屋台	0.165	ママ	0.124
お祭り	0.123	好き	0.122
(a) 夏祭り		(b) 修学旅行	

ないため、話単位でイベントの推定を行った。

今回、推定を行う対象としたイベントは、イベント特徴語辞書の作成に用いた20種のイベントに含まれ、かつ表1の上位から“文化祭”, “バレンタイン”, “夏祭り”, “クリスマス”, “勉強会”, “誕生日”, “体育祭”, “修学旅行”, “ショッピング”, “初詣”のイベント10種である。

イベント特徴語辞書の作成に用いた表3の11作品の各々について、それぞれの話の中に含まれるテキストをMeCabを用いて形態素解析を行い、名詞を抽出した。イベント特徴語辞書を参照して抽出された名詞が属するイベントを特定し、重複を許してイベントごとに計数した。計数した結果をもとに作品 a ($a=0$ から W まで)の i 話 ($i=1$ から T まで)に対して10種のイベント(e_1, e_2, \dots, e_{10})を基底とする10次元のベクトル $E_a(i)$ (以下、イベントベクトル)を作成した。

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

単話からイベントを推定する手法の特性上、本来のイベントに関係のないイベント特徴語が偶発的に出現する可能性があるため、カットオフ基準 N を設けた。例えば、ホワイトデーのイベントが発生している話では「バレンタインのお返し」など他のイベント特徴語 (i.e., バレンタイン) が出現する。こうした他のイベントの特徴語は、本稿の手法ではイベント推定のノイズとなる恐れがある。イベント特徴語の出現回数がカットオフ基準以下となるイベントのベクトルの要素に0を割り当てた。このようにして得られたベクトルを要素の和が1.0となるように正規化した。

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

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

その上で、最も値が大きい要素が閾値を超えた場合にそれをイベント推定結果として採用した。

表6 イベント毎のイベント特徴語数

	文化祭	バレンタイン	夏祭り	クリスマス	勉強会	誕生日	体育祭	修学旅行	ショッピング	初詣
非洗練辞書	50	50	50	50	50	50	50	50	50	50
洗練辞書	24	12	22	10	21	9	31	14	13	27

表7 非洗練辞書を用いた全文に対するパラメータ選定

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

表8 洗練辞書を用いた全文に対するパラメータ選定

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

表9 非洗練辞書を用いたセリフ文に対するパラメータ選定

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

表10 洗練辞書を用いたセリフ文に対するパラメータ選定

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

6 実験

本章では、5章で述べたカットオフ基準 N とイベント推定閾値 p を設定するために行った予備実験と設定したパラメータを用いて行った小説とコミックそれぞれのイベント推定結果について述べる。

6.1 パラメータ設定のための予備実験

カットオフ基準 N とイベント推定閾値 p を設定するための予備実験を行った。まず、表3のイベント特徴語辞書の作成に用いた小説11作品を対象に、各話のイベントを推定した。イベントの推定には、1話に含まれる全てのテキスト（以下、全文と記す）を用いた。次に、推定精度の評価指標として再現率、適合率、F値を算出した。算出したF値が最も高くなる N と p の組み合わせをグリッドサーチにより選定した。このとき、 N は1から6までの1刻み、 p は0.30から0.60までの0.10刻みとした。

4章にて構築した非洗練辞書、洗練辞書それぞれを用いた場合の結果を各々表7、8に示す。非洗練辞書においては、 $N=1$ 、 $p=0.30$ 、洗練辞書においては $N=6$ 、 $p=0.40$ の組み合わせでF値が最大となった。小説の全文を対象としたイベント推定においては、洗練辞書を用いたイベント推定の方が0.114ポイント上回る結果となった。

次に、コミックではセリフ文がイベントの推定に用いるテキストであり、推定の手がかりとなる単語数が小説文に比べて少なくなることを考慮し、同作品のセリフ文のみを取り出して同様の予備実験を行った。表9、10にその結果を示す。非洗練辞書においては、 $N=4$ 、 $p=0.30$ 、洗練辞書においては $N=3$ 、 $p=0.30$ の組み合わせでF値が最大となった。小説のセリフ文を対象と

したイベント推定においても、洗練辞書を用いたイベント推定の方が0.090ポイント上回る結果となった。

小説の全文を対象とした場合とセリフ文のみを対象とした場合のいずれにおいても洗練辞書を用いた方が、イベント推定の精度が高くなることが確認された。これは、非洗練辞書にはイベントにあまり関係のないと考えられるような日常的に用いられる単語が含まれていたためだと推察される。例えば4.2節の表5(b)より、修学旅行イベントの辞書には“少女”、“委員長”、“先生”などの単語が含まれていた。このような日常的に使われる単語はテキストに頻出しやすく、イベント特徴語として多く抽出される。その結果、1話あたりの各イベントにおけるイベント特徴語の数が増加することとなる。そのため、日常的に使われる単語はイベント推定のノイズとなる。イベント特徴語に含まれるノイズが多いため、 N の値を大きくしてもカットオフ基準が機能しなくなる。その結果、イベントベクトルの複数の要素に値が入り、推定対象のイベントが発生していない話に対してイベントが発生したという誤推定やラベル付けされたイベントとは別のイベントを推定するといったことが起こりやすくなった。以上のことから、非洗練辞書を用いたイベント推定の精度が低くなった。

以上の結果から、本実験ではイベント特徴語辞書として洗練辞書を用いることとし、小説においては $N=6$ 、 $p=0.40$ の組み合わせを、コミックにおいては $N=3$ 、 $p=0.30$ の組み合わせを採用することとした。

6.2 本実験

6.1節にて選定した N と p を用いて、新たに小説10作品とコミック2作品のテキストを対象に、本稿で構築したイベント特徴語辞書の評価実験を行った。

表 11 本実験に用いた小説一覧

作品名 (URL)	作者名
愛しの眠り …… 姫?! (https://ncode.syosetu.com/n3483hs/)**	平遊
幼なじみが俺の青春を全力で潰しにくるのですが? (https://ncode.syosetu.com/n9803ho/)*	黒百合咲夜
君の夢を見る (https://ncode.syosetu.com/n6500hi/)*	—
高校からやり直し！幼馴染として逆行付添人の天音ちゃんは君をとにかく幸せにしたい (https://ncode.syosetu.com/n8985hf/)**	江戸 清水
十七歳は御多忙申し上げます (https://ncode.syosetu.com/n3682fj/)**	香月よう子
小説投稿サイトを利用していたら、クラスの美少女《ヒロイン》が読者だった (https://ncode.syosetu.com/n6893en/)**	雨ノ日秋作
どうやら俺をいじめているギャルが俺のことを好きらしい (https://ncode.syosetu.com/n0223hd/)**	べいべい
花田種苗の 5 人姉弟妹 (https://ncode.syosetu.com/n0237fj/)**	グレープヒヤシンス
ヤンデレの皮を被った女の子。 (https://ncode.syosetu.com/n3156gx/)**	りぶろ
私のことを好きで好きでたまらない娘を言い当てたら 10 億円 (https://ncode.syosetu.com/n5685ho/)**	三色ライト

* (2022/10/1 存在確認) ** (2023/2/24 存在確認)

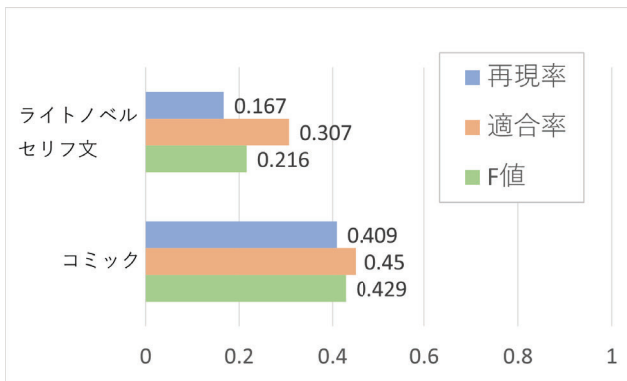


図 2 イベント推定結果の比較

6.2.1 小説

本稿で構築したイベント特徴語辞書を用いたイベント推定が可能であるか検証を行うため、イベント特徴語辞書の構築に用いなかった小説を対象に実験を行った。本実験に用いた小説 10 作品を表 11 に示す。

小説 10 作品における再現率は 0.465、適合率は 0.355、F 値は 0.403 となった。この結果から、提案手法により小説に対して一定程度のイベント推定が可能であることが示唆された。

6.2.2 コミック

本稿では、「好きな子がめがねを忘れた」[5]と「男子高校生の日常」[6]の 2 作品のコミックを対象として、6.1 節にて決定した $N = 2$, $p = 0.60$ の条件下において、イベント特徴語辞書を用いたイベント推定の精度評価実験を行った。また、6.2.1 節で用いた同小説 10 作品のセリフ文のみを対象としたイベント推定結果と比較を行った。その結果を図 2 に示す。

コミック 2 作品における再現率は 0.409、適合率は

0.450、F 値は 0.429 であった。この結果から、提案手法によりコミックに対しても一定程度のイベント推定が可能であることが示唆された。また、表 12 にコミック 2 作品で発生したそれぞれのイベントにおけるイベント推定の結果を示す。

6.2.3 考察

図 2 より、コミックは小説のセリフ文に比べて再現率、適合率、F 値の全ての値が上回る結果となった。これは、コミックにはしばしば説明セリフと呼ばれるセリフで状況を説明することがあり、小説のセリフ文と比較してイベント特徴語を多く含むためと推察される。

表 12 より、バレンタイン、体育祭のイベントについては、イベントの推定精度が高かった。正しく推定された話ではイベント特徴語が多く出現していたことが確認された。一方、クリスマス (2 話)、誕生日 (2 話) のイベントについては、推定されなかった。これは、どちらのイベントもギャグ漫画である「男子高校生の日常」で発生しており、クリスマス、誕生日に該当する話のいずれについてもイベント特徴語が少なかったことが原因である。また、今回対象としたコミック 2 作品では発生することのなかった勉強会 (4 話)、修学旅行 (1 話)、初詣 (1 話) が誤推定された。これは勉強会に属するイベント特徴語である“教科書”や修学旅行に属する“部屋”、初詣に属する“お願い”などが原因と考えられる。“教科書”は学生が日常的に使う単語であるため授業中のシーンで用いられることが多く誤推定に繋がった。“部屋”はホテルの部屋割りや部屋決めなど修学旅行に関連した単語であったためイベント特徴語として選定したが、日常会話で用いられることもあり誤推定に繋がった。

表12 コミックのそれぞれのイベントにおけるイベント推定結果

	文化祭	バレンタイン	夏祭り	クリスマス	勉強会	誕生日	体育祭	修学旅行	ショッピング	初詣
再現率	0.400	1.000	0.250	0.000	—	0.000	1.000	—	—	—
適合率	0.800	0.667	0.500	—	0.000	—	0.500	0.000	—	0.000
F 値	0.533	0.800	0.333	—	—	—	0.667	—	—	—

“お願い”は『「何お願いしたの?」「秘密だよ。言うとならないらしいから」』⁹や『私たちの番がやってきたので、お賽銭を投げて二礼二拍。心の中でお願いをして、一礼。』¹⁰のようにイベントで出現しやすい単語をイベント特徴語として選定したため、人に頼み事をするとといった特定の話の誤推定してしまった。これらのことを踏まえて、他のイベント特徴語との共起関係を利用することによりイベント推定の精度が向上すると考える。

7 おわりに

本稿では、学園ジャンルのライトノベルからテキストを収集し、イベント特徴語を抽出して構築したイベント特徴語辞書を活用することで学園ジャンルのコミックのイベントの推定を試みた。まず、学園ジャンルで発生するイベントを整理することにより、本研究の対象とするイベントを明らかにした。予備実験の結果から、イベント特徴語辞書にイベントに関係の薄い単語が含まれることにより、イベント推定の精度が下がることが確認された。本実験の結果から、ライトノベルから得られたイベント特徴語を用いる提案手法によるイベント推定を行うことで、コミックのイベント推定が可能であることが確認された。今後は、イベント特徴語辞書の拡充をすすめ、イベント特徴語間の共起関係を利用することにより、イベント推定の精度向上を目指す。

謝辞

実験で使用したライトノベル並びにコミックの作品著者各位に深謝する。また、本研究は科研費(22K12338)の支援のもと行われた。記して謝意を表す。

参考文献

- [1] Murai, H.: Automatic Extraction of Reversal-Type Punch Lines in Shin'ichi Hoshi's Flash Fictions, *Journal of the Japanese Association for Digital Humanities*, Vol. 2, No. 1, pp. 31-47 (2017).
- [2] 中村祥吾, 村井 源: ロールプレイングゲームにおける物語の複合的構造の時系列的变化と作品間比較, 第36回人工知能学会全国大会論文集, 1P1-GS-10-05 (2022).
- [3] 内山清子, 石川 諒: 恋愛漫画における物語構造分析とその応用, 第31回人工知能学会全国大会論文集,

⁹オタクな彼女は好きですか?, 第166話「風と人気者、初詣に行く4。」

¹⁰サボリ魔先輩に恋に落ちるまで, 第14話「後2カ月〜いつもと違う初詣〜」

1D2-OS-29a-3 (2017).

- [4] 松下光範: コミック工学—漫画を対象としたコンテンツ指向研究, システム/制御/情報, Vol. 63, No. 3, pp. 113-117 (2019).
- [5] 藤近小梅: 好きな子がめがねを忘れた, スクウェア・エニックス (2019).
- [6] 山内泰延: 男子高校生の日常, スクウェア・エニックス (2010).