

レビューに基づくコミック探索システムの評価

Evaluation of a Review-based Comic Search System

朴 炳宣^{*1} 山下 諒^{*2} 松下 光範^{*3}
 Byeongseon Park Ryo Yamashita Mitsunori Matsushita

^{*1*}関西大学 総合情報学部 ^{*2}関西大学大学院 総合情報学研究科
 Kansai University, Faculty of Informatics Graduate School of Informatics, Kansai University

The objective of our research is to develop a search system that supports the exploration of comics. When a user searches for a new comic based on his/her taste, the current system only provides search results based on bibliographic information (e.g., author, genre) and the contents of the comics are not taken into account. To develop a search system that provides results that reflect the taste of the user in a better manner, extraction of content-related information from a comic is desirable. However, it is difficult to acquire such information from all the contents of a comic because they consist of several modalities such as images, texts, and symbols. To overcome this difficulty, we built a search system that uses information acquired from the reviews of the comics. In this paper, we evaluate our system to verify its advantages and disadvantages and describe the patterns of the user's search behavior when using the system.

1. はじめに

出版指標年報によると、日本国内で出版されるコミック新刊の点数は年々増加しており、それらに対するアクセス手段も複数存在する (e.g., 電子書籍販売サイト, 漫画関連書籍検索). これらのアクセス手段を用いることで, 作品のタイトルや著者, 掲載誌などの書誌に関する情報 (以下, 書誌情報と記す) に基づいた検索が可能である. 一方で, ストーリーや登場キャラクターなどのコンテンツ内容に関する情報 (以下, 内容情報と記す) をトリガとした検索手段は, “ファンタジー” や “恋愛” などの大まかなジャンル情報を用いた検索に留まっている. これは, コミックがテキストとイラストの連携によってストーリーを構成するマルチモーダルなコンテンツであり各要素の識別が困難であるため, 自動で内容情報を獲得できないという点が要因の1つとして挙げられる.

このような背景のもと, 著者らはこれまでコミックの内容情報に基づくアクセス支援の実現を目指し, 各コミックに対するWeb上レビューを用いた間接的な内容情報の抽出, および探索インタフェースの検討を行ってきた.

本稿では提案インタフェースに対して以下の二つの観点から評価を行う.

1つ目は提案インタフェースで用いた情報の妥当性に関する評価である. 提案インタフェースは漫画レビューから情報を抽出しており, ユーザは抽出した情報に基づいてコミック探索を行うことができる. 本稿では, 提案インタフェースに用いるべき情報を比較実験によって評価を行う.

2つ目は, 提案インタフェースの有効性に関する評価である. 本稿では提案インタフェースを用いることで, 要求が曖昧なユーザが自分の嗜好に合致した作品にアクセスできたかについて評価を行う. さらに, コミック探索時のユーザのふるまいを観察し, システムの改善点を明らかにする.

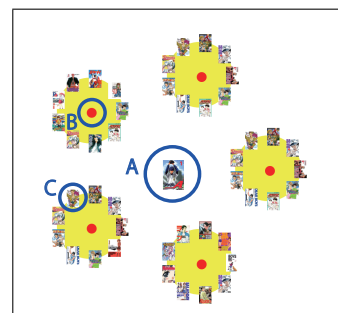


図 1: 発散的コミック検索画面

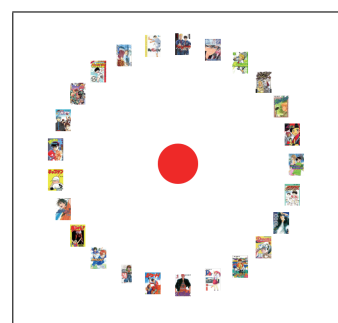


図 2: 収束的コミック検索画面

2. 先行の取り組みと課題

2.1 システムの概要

本稿で提案するシステムを図1及び図2に示す. 提案システムでは, 幅広いトピックのコミックにアクセスできる発散的な検索 (図1) とユーザが選択したトピックに関するコミックのみを閲覧できる収束的な検索 (図2) の2つの検索要求に対応することができる.

提案システムの入力クエリは「ユーザが好きなコミックタイ

連絡先: 朴 炳宣, 関西大学総合情報学部, 大阪府高槻市霊仙寺町 2-1-1, k281401@kansai-u.ac.jp

トル」である。コミックのタイトルをシステムに入力することで、図 1 の状態に遷移し、入力したコミックを起点として探索を行うことができる。

図 1 は、提案システムの主となる画面であり、3 つの情報から構成されている。1 つ目の情報は、図 1 中 A が示した箇所であり、ユーザが入力もしくは選択したコミック（以下、選択コミック）である。2 つ目の情報は、図 1 中 B が示した箇所であり、選択コミックに紐づいた特徴語と共起した語を有するトピック（以下、関連トピック）である。3 つ目の情報は、図 1 中 C が示した箇所であり、各関連トピックに含まれる語と共起した語を特徴語に持つコミック（以下、関連コミック）である。それぞれの情報にはレビューから抽出した情報が紐づいており、提案システムではそれらの語の共起を用いて“関連”を定義している。ユーザは、一度に複数のトピックに関する情報を得られるため、幅広い内容のコミックの情報にアクセスが可能である。提案システムでは、関連トピックの提示上限数を 6、関連コミックの提示上限数を 10 とし、それ以上の数が関連している場合はランダムに情報を提示している。

図 1 中 C を選択すると図 2 に遷移する。この時、ユーザが能動的に関心を持ったトピックを選択するため、複数のトピックから自身の要求に沿った情報を選択したと言える。この画面では選択したトピックの関連コミックを情報提示数に制限をかけることなく閲覧することができる。また、提示されている関連コミックのアイコンをクリックするとそのコミックが選択コミックとなって図 1 に遷移し、再度選択コミックを起点とした検索ができる。

上述したように探索行為を繰り返すことによって、ユーザは様々なコミックの情報にアクセスしながら要求に沿ったコミックを探ることが可能となった。また、既存サービスではできなかった、要求が曖昧なユーザであっても探索を繰り返すことで自分の嗜好に合致した作品へアクセスすることが可能になった。なお、詳細なデザイン指針については先行研究を参照されたい [山下 15]。

3. hLDA パラメータの選定

本稿では漫画レビューからの情報抽出手法として hLDA (hierarchical Latent Dirichlet Allocation) を用いている。hLDA とは、文書に含まれるトピックが階層構造であると仮定しながらトピック分類を行うモデルである。hLDA 法で分析を行う際には、ハイパーパラメータ (α , γ , η) と分析で考慮する階層数をあらかじめ設定する必要がある。

本実験では、複数パラメータの結果を用いた比較実験を行い、提案システムに適した情報を得るためのパラメータを選定する。本稿では、パラメータの選定にあたり (1) 分析結果の構造、(2) 各トピックに分類される語、の 2 点に着目した。

提案システムの対象コミックは、漫画レビューサイトに投稿されたレビュー数の上位作品である。そのため、対象コミックのジャンルや内容は様々であり、分類されるトピック数が多くなると考えられる。一方で、単一のコミックのみで構成されたトピックはこれに適さない。したがって、提案システムに用いる情報は、生成されるトピック数がより多く、かつ単一作品に基づいて生成トピックの数が少ない結果であるパラメータを選定することが望ましい。

また、hLDA の分析結果はパラメータ η と γ が大きく影響する [Blei 10]。一般的に、 γ が大きく η が小さい場合は、各トピックにおいて生成確率の高い単語が上位に分類されたトピックが生成され易くなる。予備実験として、各パラメータを変化

表 1: パラメータ選定実験に用いるパラメータ

ラベル	η	γ	条件 1	条件 2
トピック 1	0.025	1	153	0
トピック 2	0.1	2	77	2
トピック 3	0.5	1	44	0
トピック 4	1.0	5	37	0
トピック 5	2.0	0.1	21	0

させた場合のトピックを調査したところ、 γ が高く η が小さいと各トピックの上位に一般名詞が多く分類される傾向が観察された。一方、 γ を固定して η を大きくすると、各トピックに固有な一般名詞が多く含まれる傾向が観察された。このように、パラメータに応じて各トピックに分類される語が異なるため、提案システムを用いるユーザにとって適切なパラメータを明らかにする必要がある。

上述した点を考慮することで、コミック探索において、より有益な情報を提示することができると考えられる。

3.1 実験に用いるパラメータの選定

評価実験を行うにあたり、実験に用いるパラメータの選定を行った。

hLDA は確率分布に基づいたトピックの推定を繰り返し行うことで結果が収束し、より確度の高い結果を得ることができる。しかし、分析データに対して各パラメータが一定の値を越えた場合は収束せず、その結果の階層構造は誤ったものとなる。このような分析結果は正しく分類されたとは言い難く、本実験に用いるパラメータとして適さない。そこで、分析結果が正しく収束しなかったパラメータを選定前にあらかじめ除外することとする。

また、上述のように各トピックに生成される確率が高い単語はパラメータ η が大きく影響する。そこで本実験では実験に用いる η の値を 0.025, 0.1, 0.5, 1.0, 2.0 の 5 段階に設定した。また、各パラメータに適した γ の値は対象データや η の値に応じて異なるため、今回はそれぞれの η の値に対して 1.0 から 5.0 まで 1 刻み、10 から 100 まで 10 刻みに変化させ、計 15 種類の結果を収集した。なお、パラメータ α の値は 1.0 に固定した。

以上の手順で選ばれたパラメータを表 1 に示す。今回は、階層 3 に分類されたトピック数から単一のコミックに関するトピックの数を引いた値が一番高い結果を選定した。ここで表 1 の“条件 1”は各パラメータを用いて分析を行った時に発生した全体トピック数を、“条件 2”は発生したトピックの中に単一のコミックしか含まれなかったトピック数を示している。

3.2 実験手続き

実験協力者は著者らの研究室に所属している大学生 7 名 (男性 6 名, 女性 1 名) である。実験を行う前に、実験協力者に対して実験目的と課題内容が説明された。本実験の課題は、(1) 提示された語群 (hLDA の推定結果) に含まれるトピックの推定、(2) 推測した話題の根拠となる語の選択、であった。

本実験におけるトピック推定は自由記述であり、単語 (e.g., SF, 未来), 複合名詞 (e.g., 宇宙飛行士, 青年漫画), 短文 (e.g., リアルな SF, 宇宙飛行士になるのが夢の主人公の物語) など制約を設けなかった。

本実験で実験協力者に取り組んでもらう課題は、各パラメータの結果をそれぞれランダムに 10 トピックずつ選定した、計 50 トピックである。本実験では回答時間を設けず、すべての課題に回答した段階で実験終了とした。

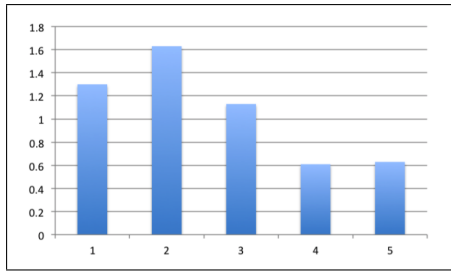


図 3: 平均回答トピック数

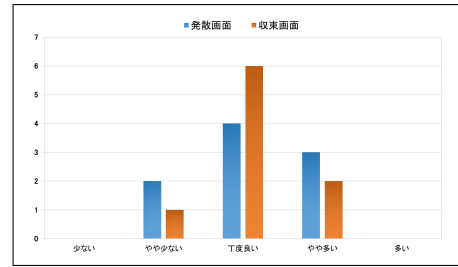


図 5: 画面内の作品数についての回答結果

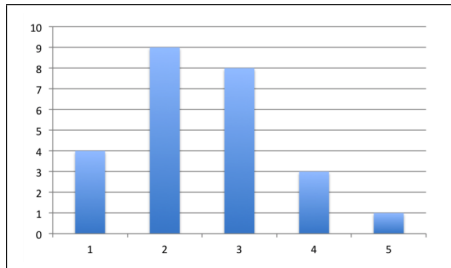


図 4: 平均共通回答トピック数

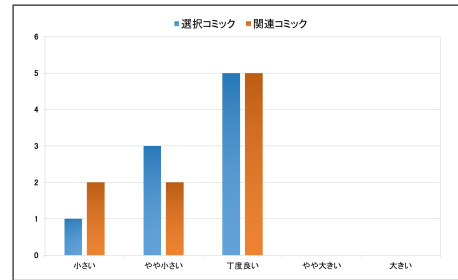


図 6: 表紙サイズについての回答結果

3.3 結果

各パラメータの全項目に対する実験協力者の回答結果の平均値を図 3 及び図 4 に示す。図 3 は各パラメータに対するトピックの平均推定数を、図 4 はトピック推定において複数の実験協力者が同一の回答を行ったトピック数をそれぞれ表している。これらの表を見るとどの指標においても、トピック 2 の結果が最も回答数が多いことが確認できる。

トピック 2 では、“麻雀”や“ギャンブル”、“パチンコ”などの一般名詞が、トピック 5 では、“フェアリーテイル”、“ベルゼブブ”などの作品タイトル名や、人名と思われる固有名詞が高い割合で含まれている。今回の実験結果が η が小さい結果の方が回答率が高かったことから、トピック推定を行うには一般名詞の方がトピックを推定し易いということが推察される。

一方で、 η が最も小さいトピック 1 (表 1 参照) ではなくトピック 2 の結果が最も高い結果となった。各回答結果を分析したところ、一般名詞だけでなく登場人物をはじめとする固有名詞を選択している回答が多く見受けられた。このことから、既知のキャラクタ名からそのキャラクタが登場する作品をイメージしてトピックの話題を推測していることが示唆された。

以上を踏まえ、提案システムでは最も回答率が多かった、 $\alpha=1.0$ 、 $\eta=0.1$ 、 $\gamma=2.0$ のパラメータで分析した結果を用いることとする。

4. ユーザレビュー

本実験の目的は、提案システムを用いてユーザ評価を行い、コミックの探索過程を観察するとともに、システムの評価を行うことである。

4.1 実験手続き

実験協力者は、関西大学総合情報学部 に在籍している学生 9 名 (男性 5 名、女性 4 名) である。本実験では、実験協力者に対して、(1) 実験目的、(2) 提案システムの概要および操作、(3) 実験課題、の 3 つの項目を説明した。

提案システムの概要説明では、まずシステムに用いている情報 (e.g., 選択コミック、関連トピック、関連コミック) を説明

し、それぞれの情報の位置づけを伝えた。また、操作説明ではクリック処理 (e.g., 関連コミックをクリックした際の提示画面の変化) と、紐づいた情報を閲覧することができるマウスオーバー処理を伝え、全ての実験協力者が課題に用いるシステムの機能を理解している状態にした。

本実験の課題は、「“面白そうな漫画”の発見」である。この“面白そう”というフレーズに対して解釈の制限を設けておらず、「面白そうな特徴語が紐づいていた」や「絵柄が好き」などといった具体的な例を交えながら自由に考えて良いことを伝えた。この課題は、実験協力者が“面白そうな漫画”を発見した、あるいは探索時間が 15 分を経過した時点で終了とした。

また、本実験では発話思考法を採用しており、実験協力者の探索時の思考を記録した。そのため、提案システムを操作する際に考えたことや感じたことを発言するように促した。なお、本実験は、提案システムに用いた作品群の中から実験協力者が好きな作品を選択してもらい、その作品が選択コミックとして提示された段階から開始とした。

さらに本実験終了後、アンケートに回答してもらった。アンケートの設問項目は (1) 課題、(2) システムデザイン、(3) システムに用いた情報、の 3 点に関するものである。

4.2 実験結果と議論

以下で、各実験協力者の探索過程の特徴とアンケート結果と課題についてそれぞれ述べる。

1. 課題

課題に関する設問に対しては、すべての実験協力者が Yes と回答した。加えて、自由記述欄から「自分の好きなジャンルを追及するのに良いシステムだと思った」や「漫画を普段読まない人に対しても便利なシステムになると期待できる」などの意見が得られた。これらの結果から、提案システムを用いることで自分の嗜好に合致した作品へのアクセスが可能であることが示唆された。

2. システムデザイン

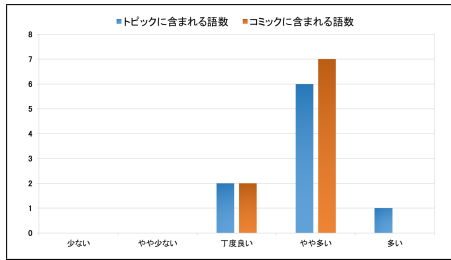


図 7: システムに用いる情報についての回答結果

システムデザインについての回答結果を図 5 と図 6 に示す。図 5 はシステムで提示される作品数の印象に対する結果であり、図 6 はシステムで提示した作品の表紙サイズに対する結果である。

システムで提示される作品数に関しては、“丁度良い”という回答が一番多いものの、“やや少ない”や“やや多い”などの不満を持った意見が半数を超え、「絵をはっきり見ることができなかった」などの回答も得られた。

システムで提示した作品の表紙サイズに関しては、どの結果においても“丁度良い”の回答数が半数を超えたものの、“やや小さい”、“小さい”を選択した回答も確認された。自由記述欄では、「作品同士の間隔が狭い」や「絵柄をよく見ようとすると小さく感じた」といった絵柄が小さいという意見が多かった。

提案システムにおける情報提示方法は、コミックが表紙情報であり、トピックが赤丸である。これらの情報はマウスオーバーすることで紐づいた情報を確認することができ、さらにコミックの場合は提示された表紙情報も拡大される。探索は提示された情報に対して取捨選択を行う行為であるため、提示される情報の可読性は高い方が望ましい。しかし、「どのような語が紐づいているのかをその度確認しなければならない」など可読性に関する意見が複数確認された。より円滑な情報アクセスを実現するために、今後は情報提示方法を改めて検討する必要があると考える。

3. システムに用いる情報

レビューから抽出したトピックの語の数に対する印象と、レビューから抽出した各コミックの特徴を表す語の数に対する印象に関する結果を図 7 に示す。

トピックの語の数に対する印象に関しては、“やや多い”という回答が最も多かった。「提示される語数が多く、詳細にどんな感じのトピックであるのか知ることができたがその分時間を要した」などの回答が見受けられ、提示された語からそのトピックを推測する際に時間を要したことが確認された。さらに、「どのような基準で関連トピックが作成されているかわからない」という回答があり、提示コミックとの関連を測ることができなかったことがわかった。

各コミックの特徴を表す語の数に対する印象に関しては、“やや多い”という回答が最も多かった。「詳細にどんな感じの漫画なのかはわかるが、もう少し大まかな内容を短時間でつかみたいと思った」や「よくわからない単語があったり、少し多すぎて見づらい気がした」などといった回答が見受けられた。

提案システムに用いた各コミックの特徴語は 50 語、各トピックの語として hLDA を用いて分類された各トピックに紐づく語を 30 語と設定し、各情報の特徴をより明確にすることを狙った。これにより、「語群から各情報の内容や話題を把握することができた」という意見を得られた一方で、「情報量が多く内容を把握することが難しかった」、「把握するために要した時間が長かった」などの意見も見受けられた。各情報の把握が遅れる場合は可読性が低くなるため、探索行為の負担になる可能性が生じる。この課題を解消するためには、少ない情報量で各情報の特徴を判別することができるような情報を紐付ける必要がある。

本稿ではレビューからの情報抽出手法として tf-idf と hLDA を用いた。tf-idf は、単語頻度と文書頻度に基づいて各文書に含まれる単語の特徴量を算出する手法であるため、分析対象の文書構造に依ることなく分析を行うことができる。しかしながら、hLDA は確率モデルに基づいて文書中のトピックを推定するため、この結果は分析時に設定するパラメータと対象の文書の構造が大きく影響する。そのため、分析対象であるレビューから適切なトピックを抽出するためにはその特徴を捉え、分析対象に適した事前処理を行うことが望ましい。

5. おわりに

本稿では、コミックの内容情報に基づいた探索的な情報アクセス支援の実現に向けて、コミック探索システムに対してユーザレビューを行った。まず提案システムは漫画レビューからの情報抽出手法として hLDA を用いており、本システムに適した hLDA のパラメータ選定するため比較実験を行った。これにより、ユーザの思考に近いトピックの分類を行っていると思われるパラメータを選定することができた。また、提案システムを用いてユーザ評価を行い、コミックの探索過程を観察するとともに、システムの評価を行った。その結果、提案システムを用いることで自分の嗜好に合致した作品へのアクセスが可能であることが明らかになった。一方、情報提示方法やシステムに用いる情報に関しては、より検討を行うべき部分が存在していることも明らかになった。特に、現状ではレビューに対して品詞に着目した情報抽出にすぎず、レビューの特徴を詳細に把握した上での分析には至っていない。そのため、レビューから多角的な情報をそれぞれ獲得するためには記述パターンを分析する必要がある。今後の展望として、ユーザの思考に則したコミック探索の実現に向けて、レビューから抽出する情報をシステムに適したものにすべく、情報源である漫画レビューの特徴を捉えていきたい。さらに、レビューの特徴を明らかにした上で、登場キャラクターと作品名の判別方法などを検討していく。

参考文献

- [山下 15] 山下諒, 松下光範: 階層的トピック分類を用いた内容情報に基づくコミック探索システムの提案, HCG シンポジウム 2015 論文集, pp.154-160 (2015).
- [Blei 10] Blei, D. M., Griffiths, T. L. and Jordan, M. I. : The Nested Chinese Restaurant Process and Bayesian Nonparametric Inference of Topic Hierarchies, Journal of the ACM, Vol. 57(2), No. 7, pp. 1-30 (2010).