

興味喚起度 MAP を利用した電子書籍小説「立ち読み」インタフェースの実装と評価

村井 聡一[†] 牛尼 剛聡^{††}

[†] 九州大学大学院芸術工学府 〒 815-8540 福岡市南区塩原 4-9-1

^{††} 九州大学大学院芸術工学研究院 〒 815-8540 福岡市南区塩原 4-9-1

E-mail: [†]2ds11094e@s.kyushu-u.ac.jp, ^{††}ushiana@design.kyushu-u.ac.jp

あらまし 書籍選別において、「立ち読み」は有効な手段とされている。しかし、Web における「立ち読み」では、閲覧可能ページが指定されていることが多く、ユーザが書籍全体から選別に必要な情報を得ることが困難である。ユーザの検索クエリに応じて提示ページを変える手法も利用されているが、未読の小説においては、小説本文に含まれるクエリを思い付くことは困難である。我々は、電子小説を対象とし、「立ち読み」を始める箇所の効率的発見を支援するインタフェースの開発を行ってきた。本インタフェースでは、小説の選別においては、興味をひかれるかどうかが重要と考え、「物語の進行における興味をひく度合いの推移」を可視化した興味喚起度 MAP を基盤とする。本論文では、我々が開発したインタフェースの実装について述べる。さらに、被験者実験の結果に基づいて、提案インタフェースの有効性を評価する。

キーワード 電子書籍, レビュー, インタフェース, 興味喚起, 情報可視化

Implementation and Evaluation of A Browsing Interface for A Novel Using Attractiveness MAP

Soichi MURAI[†] and Taketoshi USHIAMA^{††}

[†] Graduate School of Design, Kyushu University
4-9-1 Shiobaru, Minami-Ku, Fukuoka 815-8540 Japan

^{††} Faculty of Design, Kyushu University
4-9-1 Shiobaru, Minami-Ku, Fukuoka 815-8540 Japan

E-mail: [†]2ds11094e@s.kyushu-u.ac.jp, ^{††}ushiana@design.kyushu-u.ac.jp

1. はじめに

近年、書籍の電子化が進み、ユーザは Web 上で書籍を探し、読むことができるようになった。電子書籍の増加にしたがって、読む書籍を見つけることが困難になり、書籍の選別を支援するシステムが重要になってきている。Web 上において、ユーザが読む書籍を発見するタスクは、「候補を絞り込む段階」と「候補の価値を判別する段階」の二つの段階に大別できる。

ユーザが Web 上に存在する全ての書籍に目を通すことは困難であるため、まずはキーワードによる検索や推薦によって、候補を絞る作業が行われる。これまで、読む書籍の候補を絞り込むことを目的とした検索や推薦に関する研究が活発に行われてきた。しかし、一般的に、提示された候補の書籍を読む価値のある書籍と、即座に断定するのではなく、何らかの方法で価

値を判断し、選別を行う。我々は、有効な選別方法の一つに、電子書籍の選別支援においてはあまり注目されてこなかったが、実際に書籍の本文を読む「立ち読み」があると考えている。

書籍等デジタル化推進事業の調査報告書 [1] によると、実世界の書店、ネット書店の両方において、書籍、コミックを購入する際に参考に行っている情報として、立ち読みをあげた人の割合が最も高くなっている。しかし、実際に立ち読みを行う人の割合においては、実世界の書店では約 40% の人が立ち読みを行なっているのに対し、インターネット書店では立ち読みを「全くしない」人の割合が 51.9% となっている。このことから、Web においても「立ち読み」は有効的だが、Web 提供されている「立ち読み」機構はユーザに対する訴求力が弱いために、実際に行う人が少なくなっていると考えられる。

実世界での立ち読みと比較し、web 上での「立ち読み」が行

られない理由として、閲覧ページの制限が考えられる。紙媒体の書籍では、好きなページを立ち読みできることが多い。一方、Web上で提供されている電子書籍の「立ち読み」機構では、ユーザが閲覧可能な範囲が先頭ページから、十ページから数十ページに限定されていることが多い。そのため、書籍にユーザの選別に役に立つ内容が書いてあったとしても、立ち読み時に閲覧できないために、読みたい書籍と判断できないことがある。また、立ち読みの魅力としてあげられる、ページを高速にめくることによる偶発的な出会い、も起きにくいと考えられる。

この問題を解決するために、制限されたページ数内で、ユーザにとって有益な文章を優先的に提示する手法が利用されている。電子書籍の「立ち読み」が利用できるサービスの一つである Google ブックス^(注1) [2] では、閲覧可能なページ数が制限されており、ユーザが与えたクエリに応じて、提示するページを変更する。クエリによる「立ち読み」可能ページの提示は、専門書や実用書等を選別するような、ユーザが知りたい事柄があらかじめ把握している状況では有効的である。科学技術論文や専門書に対しては、効率的に文書の内容を把握するための、概要の生成や、閲覧支援に関する研究も行われてきた [3], [4]。しかし、電子書籍における大きなジャンルの一つである小説に対しては、科学技術論文や専門書を対象とした文書検索機構や、閲覧支援機構が不適切である場合が多い。これは、未読小説の選別をする場合、小説本文に含まれる単語が未知であるために、ユーザは興味のあるキーワードの想起が困難であることが原因である。

以上の背景より、我々は書籍の中でも小説を対象を限定し、「立ち読み」の支援を行うインタフェースを開発した。我々は、未読小説の選別においては、興味をひかれる箇所を発見できるかどうかを重要と考え、「物語の進行」に基づいた「箇所がユーザの興味をひく度合い」の変化を可視化した「興味喚起度 MAP」を利用した「立ち読み」を提案する。

本論文の構成は以下の通りである。第2章で関連研究を述べる。第3章で提案するインタフェースの概要について述べる。第4章で提案するインタフェースを実現するための手法を簡述する。第5章で提案するインタフェースでの有効性を測る被験者実験について述べ、第6章でその結果の考察を行う。第7章で本稿のまとめについて述べる。

2. 関連研究

我々の研究と関わる分野は大きく、情報抽出分野と情報可視化分野とに分けられる。この章では、最初に、書籍選別における「立ち読み」に関する先行研究について述べる。次に、情報抽出分野と情報可視化分野における関連研究について述べる。

2.1 「立ち読み」に関する研究

まず、本研究でいう「立ち読み」を、松田 [5] の研究をもとに定義する。松田は、図書館と書店における行動観察を行った結果、「ブラウジング」を『出会わなければ必要か判断できない情報を含む、曖昧さを持つ情報欲求を満たすため、何らかの期

待を抱きながら、利用できる感覚全てを用いて広範で多量な情報源から何らかの基準で必要なものを選び取る情報獲得の一手法である』と定義している。本研究では、「立ち読み」を『書籍を選別するために、書籍の本文を「ブラウジング」する行為』、効率的な「立ち読み」を『ユーザにとって、読みたいと思う書籍かどうかを、少ない時間、少ない文章量で判断すること』と定義する。

書籍選別に関する研究の中で、書籍本文をブラウジングする行為に関する研究について述べる。Annika [6] らは、学術的、調査目的に使われる図書館の利用者を観察し、彼らにインタビューを行った。Annika は、手にとった書籍を、必要な書籍と判断するのに、目次、読書、フリッキングが役立っていたと報告している。松田も書籍を適当に開いて飛ばしながら見るという行為を、最も頻繁に観察された行為の一つとして報告している。しかしながら、両研究では、行為の対象を小説に限定していない。

小説の選別に限定した研究には、Ooi [7] の研究があるが、「立ち読み」に関する記述はあまりない。また、電子書籍においては、Dana [8] は学術的な内容の書籍を対象にした、トランジションログの解析を行っている。

電子書籍の小説に対する「立ち読み」の有効性は未知数であるが、我々は、実世界で有効である「立ち読み」は、Webにおいても同様に有効だと考える。Webにおける「立ち読み」においても、漠然とした興味がある段階では、小説を冒頭から読むのではなく、小説全体を俯瞰することで、興味を喚起する内容を発見することが重要だと考え、小説全体を俯瞰できる興味喚起度 MAP を提案した。

2.2 情報抽出に関する研究

ユーザの興味喚起に関する情報を抽出する研究としては、沢井ら [9] の研究があげられる。沢井らは、大衆の興味が順位情報付き文書に反映していると仮定し、まず素性（単語）の興味の強弱を求め、文章に出現する素性の興味スコアの平均によって文章興味値スコアを求める手法を提案している。沢井らの研究において、注目する興味とは、語句そのものが普遍的に持つ興味であり、例として、高い興味スコアを持つ語として、「怪死」、「偏愛」、低い興味スコアを持つ語として、「会社」、「東京」等が抽出される。

我々は、普遍的に興味をひくような単語でも、小説の特徴を表していなければならないと考えている。また逆に、一般的には興味を引きそうにない単語でも、小説において特徴的、印象的な使われ方をしている単語は、興味をひく単語だと考える。本論文で提案するインタフェースでは、興味を喚起すると思われる単語の興味喚起度を、小説のレビューから計算し、単語の興味喚起度に基づいて、小説内のある文脈範囲の興味喚起度を導出する。本インタフェースは、ユーザは提示されたキーワードの中から、自分に興味がある対象を選択可能である。

2.3 情報可視化に関する研究

2.3.1 文学作品の可視化に関する研究

谷口 [10] は小説を深く鑑賞するために、三島由紀夫著『豊饒の海』の四部作それぞれに対する小説構造を、登場人物と、

(注1) : <http://books.google.co.jp/>

キーワードを利用した等高線グラフの文章地図による可視化を行っている。対象とする小説に詳しい人が文章地図を見ると、ある箇所がどのような場面なのかがわかるようになっている。しかし、未読者にとっては、どのような場面であるかを推測することは難しい。また、表示する単語を小説本文に出現する頻度の高い単語の中から、ヒューリスティックを用いて抽出しているが、科学技術論文等と異なり、小説においては本文における出現頻度が高い単語が、本文の首題を適切に表しているとは限らない。山田 [11] の、シェイクスピアのストーリーの可視化、Plaisant ら [12] の著者の年代毎の特徴の可視化があるが、これらもインタフェースを使うユーザを、対象書籍に詳しい人物としており、人手による単語の選択や重み付けを必要としている。未読者であるユーザが「立ち読み」を行う際に、適切な単語を指定することは困難である。また、Web 上にある全ての書籍に対して、あらかじめ人手で重要な単語と、その重要度を決定するのは現実的ではない。

「立ち読み」において、ユーザは小説の深い読み解きを行わず、ざっと読むことで、自分が読みたい小説かどうかを判断することを行うと考えられる。我々は、Web 上にあるユーザのレビューを利用することにより、小説の主題に関わる単語から、読者の興味を喚起する可能性が高い単語を高精度に抽出し、それを基に小説の興味をひく度合いの推移を可視化する。

2.3.2 ドキュメントの可視化に関する研究

対象とする文章、文章群の内容を把握するために、その特徴的な単語を示すことは有効的である。ドキュメント中で、重要な単語を提示する方法に、タグクラウドがある。タグクラウドの問題点として、ドキュメント中における単語の出現位置がわからないことがあげられる。この問題に対する提案として、タグ中に単語の出現頻度の推移がわかる情報を加えたものが、幾つか提案された [13], [14]。

我々が提案するインタフェースは、操作対象を小説に限定している。小説において、単語がどの位置に出現するのか、という情報は、読み始める箇所を探す手がかりとなる。しかしながら、小説本文においては、数回しか出現しない単語も多い。そのため、各単語の出現位置を示すのではなく、小説の物語の進行を横軸にとった、一つの大きなチャートをつくり、随所に特徴的な単語を配置させる方が好ましいと考えた。我々の提案する興味喚起度 MAP に近い表現に、Havre ら [15] の ThemeRiver があげられる。ThemeRiver は、横軸に時間の進行、縦軸に重要度を取り、層の中に各時点のキーワードを配置する代表的な方法である。ThemeRiver は、現在に至るまでに、様々なドキュメントにおいて応用されている。例えば、Liu ら [16] は、電子メールのアーカイブを対象に、メール中に出現する単語頻度の時系列的な変化を可視化したインタフェースを提案している。このインタフェースでは、時間を横軸にとり、縦軸上にトピックの重要度が表現されている。そして、各層のトピックを構成する単語を、タグクラウドで表示することで、電子メールアーカイブの話題の全体像を把握し、ユーザが興味のある内容のメールを効率的に発見し閲覧することを可能としている。本論文で提案するインタフェースでは、興味喚起度 MAP により、

ユーザが小説の全体的な流れ俯瞰し、興味を持った箇所を閲覧可能とする。

3. 「立ち読み」インタフェース

我々が開発した「立ち読み」インタフェースは、興味喚起度 MAP を基盤としている。興味喚起度 MAP は、「物語の進行」に伴って、「ユーザの興味を喚起する度合い」の推移をグラフで可視化し、グラフ上の特徴的な箇所にユーザの興味を喚起する単語を配置した図である。興味喚起度 MAP を利用することで、ユーザが興味をひく可能性が高い箇所を、「立ち読み」を開始する箇所として効率的に選択することを支援するインタフェースの開発を目指す。

図 1 は、太宰治著「斜陽」を提案インタフェースのスナップショットである。本「立ち読み」インタフェースは、web ブラウザ上で動作する。表示画面は、俯瞰画面、ズーム画面、立ち読み画面からなる。俯瞰画面とズーム画面は processing.js によって開発した。

図 1 の興味喚起度 MAP の俯瞰画面は、小説全体の流れを把握するために用いる。興味喚起度 MAP の横軸は物語の進行を表し、縦軸は特定の箇所がユーザの興味を引く度合いの大きさを表している。層がピークとなる箇所には、その箇所において特徴的であり、興味を引くと予想される単語が配置される。ズーム画面は、ユーザが俯瞰画面において注目しているピークを拡大して表示する。立ち読み画面は、ユーザが注目する箇所の文章を表示する。

ユーザはホイール付きの 2 ボタンマウスによって、「立ち読み」インタフェースを操作する。操作の概要を以下に示す。

- 左クリック：俯瞰画面とズーム画面において、ユーザが単語を左クリックすると、立ち読み画面に、選択した単語を中心とした文章が表示され、ユーザは「立ち読み」を行うことができる。また、層内で単語以外の箇所を左クリックした場合は、クリックした箇所にあたる形態素を中心とした文章が表示される。例えば、図 1 は、単語「戦闘」を左クリックした状況の様子を表しており、立ち読み画面の文章には選択した「戦闘」を含む文章が表示される。

- ホイールクリック：適切な興味喚起度の推移の粒度は、「立ち読み」の状況によって異なると考えられる。例えば、ユーザが「立ち読み」を開始してから、しばらくの間は、全体を広く俯瞰するために、大きな粒度で小説の特徴を把握できることが望ましい。一方、ユーザが興味を持つおおよその位置を見つけたときには、より細かい粒度で文章の特徴を把握することが必要になる。そこで、本インタフェースでは、小説中の興味喚起度を計算するために考慮する文脈の範囲の粒度を複数用意し、ホイールクリックにより切り替え可能とした。具体的には、上位の層が荒い粒度での興味喚起度を表し、下位の層は粒度が細くなるようになっている。この機能を用いて、上層部の層で興味をひく箇所に目星をつけ、その箇所粒度を細かくして表示することにより、「立ち読み」を開始する箇所を絞り込むことを可能とする。

例えば、図 2 では、ユーザが「戦闘」が出現する箇所に興味

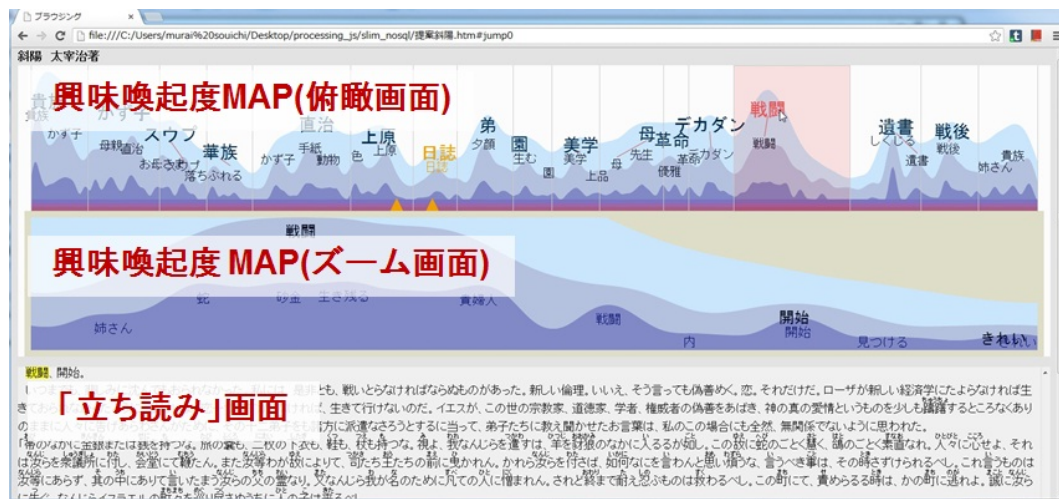


図1 提案インタフェースによる『斜陽』の表示例

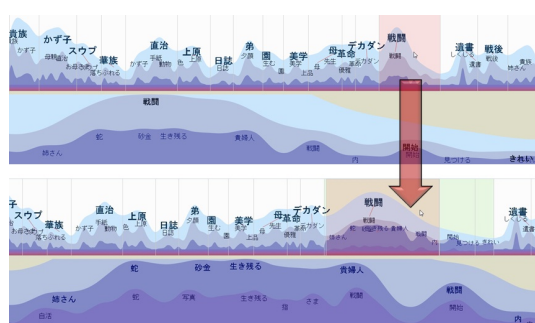


図2 注目範囲のより細かい粒度の層を表示した例

を持ち、その箇所についてももう少し詳しく知りたい時に、層を一段階下げる（粒度を一段細かくする）ことで、詳細なレベルでの「生き残る」「貴婦人」といった単語が提示される。

- 右クリック：「立ち読み」を進めると、ある単語に興味を持ち、その単語が含まれる前後の文章を読みたい、という要求が生まれることがある。そこで、俯瞰画面、またはズーム画面に表示されている単語を右クリックすると、小説中からその単語が含まれる箇所を検索し、その単語が出現する位置を、俯瞰画面上に三角形で表示する。さらに、三角形を左クリックすることで、その位置を中心とした文章の一部分を、立ち読み画面に提示する。例えば、図1は、単語「日誌」が右クリックされた状況であり、「日誌」が小説のどこに出現するかを閲覧画面の下部の三角形の位置によって把握することができる。小説においては、未読状態では適切なクエリを思い付くことが難しく、「立ち読み」を進めるうちに検索クエリが見つかると考え、クエリをキーボードで入力するのではなく、右クリックで選択するだけで検索可能とした。

4. 処理の流れ

本論文で提案するインタフェースにおける処理の流れを、図3に示す。

本インタフェースでは、最初に、小説本文に出現する単語の興味喚起度を、Web上に存在する小説のレビューを使って決定

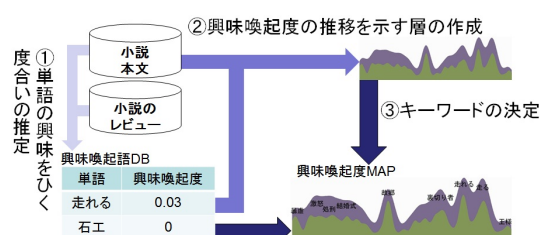


図3 処理の概要

する。対象とする小説のレビューに出現する頻度が高く、他の小説のレビューの出現頻度が低い単語に高い興味喚起度を与える。次に、興味喚起度MAPを作成する。興味喚起度MAPでは、高い興味喚起度が密に出現する箇所が、高い興味喚起度をとるように可視化する。これにより、ユーザは興味を喚起する文章に出会う可能性が高い箇所を知ることができる。しかし、そこにどのような内容に関する文章が書かれているかは分からない。そこで、それぞれの層において、上に凸となる区間において特徴的な単語を表示する。最後に、生成された興味喚起度MAPを利用して「立ち読み」に必要な機能を提供するインタフェースとしてユーザに提示する。

4.1 単語の興味喚起度

本手法では、Web上のレビューを利用して、電子書籍小説に出現する単語の興味喚起度を推定する。レビューには、読者が書籍を読んだ感想が含まれており、読書中に興味を喚起された場面や人物に関する記述が含まれている。したがって、小説が読者に与えた印象という点から、小説の特徴を抽出できると考え、興味喚起度を推定する。

まず、小説本文に出現する単語を、そのレビューにおける出現に基づいて、以下の2種類のカテゴリに分類する。

- 小説本文にもそのレビューにも出現する単語 (カテゴリ A)
- 小説本文には出現するが、そのレビューには出現しない単語 (カテゴリ B)

ここで、カテゴリAは小説の中で、レビューの興味を喚起した事項に関連する単語が含まれている可能性が高い。一方、

カテゴリ B の単語は、レビューがレビューで言及しなかったことから、興味とは直接的に関係しない可能性が高い単語と考えられる。我々は、「レビューが興味を感じた対象は、他の読者の興味を引く可能性が高い」と仮定し、カテゴリ A に含まれる単語がユーザーの興味を喚起する度合い（興味喚起度）を求める。なお、カテゴリ A 以外に含まれる単語の興味喚起度は 0 とする。ここで対象とする単語は、名詞（代名詞、接尾、数を除く）、動詞（自立動詞以外を除く）とする。ただし、他の語句の補助的な機能を持つ語句であることから、興味の強弱の判断材料として不向きであるため「する」、「ある」、「よる」、「いる」、「なる」、「いう」、「みる」、「できる」の 8 形態素 [9]、平仮名一文字、カタカナ一文字はストップワードとする。

我々の過去に行った実験結果により、小説本文をドキュメント単位とした際の TF-IDF 値よりも、有意に興味をひく単語を抽出できることが示されている [17]。また、登場人物名は、未読者の興味をひかないことが、我々の過去の被験者実験により示されているので、人物名の興味喚起度は 0 とする。

4.1.1 興味喚起度の推定手法

対象とする小説とそのレビューの両方に含まれる単語が、ユーザーの興味を喚起する可能性は同一ではない。本手法では、興味喚起度を小説における出現頻度ではなく、レビューにおける出現頻度に基づいて求める。具体的には、対象とする小説のレビューに出現する頻度が高い単語は、読者の興味を引いた度合いが高いとし、興味喚起度として高い値を与える。また、他の小説のレビューにも頻出する単語は、レビューに一般的に利用される概念であり、小説に関する可能性は低いと考える。

書籍集合を B 、小説 b に対するレビューの集合を $R(b)$ とする。注目する小説を b としたとき、単語 t_i の興味喚起度 at を以下の式で定義する。

$$at(t_i, b) = \sum_{r \in R(b)} tf(t_i, r) \cdot \log \frac{|B|}{|\{b : r \in R(b), r \ni t_i\}|} \quad (1)$$

4.2 小説の特定箇所の興味喚起度とその変化

提案インタフェースでは、小説の興味喚起度の変化を提示することにより、ユーザーが興味を持つ可能性が高い箇所を把握し、効率的に閲覧可能とする。本インタフェースでは、小説を構成する最小の意味的なまとまりは、小説中の形態素であると考え、「興味を喚起する単語を含む文脈は、興味を喚起する可能性が高い」と仮定し、高い興味喚起度をもつ単語が密に出現する箇所は、小説として興味喚起度が高いと考える。形式的には、小説に出現する形態素の出現位置を先頭から順番に i で表すとき、形態素出現位置 i を中心とした文脈の興味喚起度を、 $i - w \leq i, i \leq i + w$ の範囲に出現する単語の興味喚起度の重み付き和として求める。本手法では、形態素出現位置 i から離れた形態素の重要度を軽く扱うためにハニング窓関数 [18] を各単語の興味喚起度に乗じた。形式的には、形態素出現位置 i の文脈範囲 w の興味喚起度 $AT(i, w)$ を以下の式で定義する。

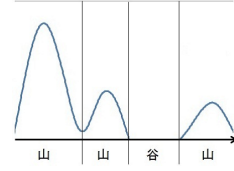


図4 「山」と「谷」

$$AT(i, w) = \sum_{c=i-w}^{i+w} at(i) \cdot \frac{1}{2} \left(1 + \cos 2\pi \frac{|c-i|}{2w} \right) \quad (2)$$

ここで、 $at(i)$ は形態素出現位置 i の単語の興味喚起度を示す。

提案インタフェースでは、物語進行に伴う興味喚起度の変化を層として表現する。層の推移の粒度は、ハニング窓の幅により決定する。ハニング窓の幅が多いときは大局的な興味喚起度の変化を表す層、ハニング窓の幅が小さいときは局所的な興味喚起度の変化を表す層となる。提案インタフェースでは、一番下層の窓幅を前後 100 語 ($w = 50$) に決定した。そして、窓幅を前後 200, 400, 800, 1600, 3200 の 6 段階を利用可能とした。しかし、小説本文の文章量が多く、層の推移が細かすぎる場合は、一番下層の窓幅を前後 200, 400, 800, 1600, 3200, 6400 の 6 段階を利用可能とした。

4.3 ユーザに提示する特徴的な単語の決定と配置

上記のように、本論文で提案するインタフェースでは、物語進行に伴う興味喚起度の変化を層として可視化することで、ユーザーは興味を喚起する文章に出会う可能性が高い箇所を知ることができる。しかし、そこにどのような内容に関する文章が書かれているかは分からない。そこで、それぞれの層において、各ピークに特徴的な単語を表示する。本論文では、上に凸となる区間を「山」と呼び、興味喚起度が 0 となる区間を「谷」と呼ぶ。図 4 に「山」と「谷」の例を示す。

「山」に特徴的な単語は、その区間において出現する単語の興味喚起度と、注目する「山」における出現頻度を考慮して決定する。具体的には、注目する層における全ての「山」からなる集合を M とし、全ての「谷」から構成される集合を V とするとき、ある「山」 m_j における単語 t_i の特徴量 un を以下の式で計算する。

$$un(t_i, m_j, w) = \left(\frac{at(t_i, w)}{\max_{t \in T} (at(t, w))} + \frac{tf(t_i, m_j)}{\sum_k tf(t_k, m_j)} \right) \cdot \log \frac{|M| + |V|}{|\{m : m \ni t_i, m \in M\}| + |\{v : v \ni t_i, v \in V\}|} \quad (3)$$

この特徴量 un は、小説の興味喚起度に基づいて、小説をいくつかの部分文書（「山」と「谷」）に分割し、それらの部分文書における単語の出現頻度に基づく局所的な重要度と、部分文書頻度の逆数に基づく大局的な重要度の積とした、単語の部分文書における重要度と解釈できる。

それぞれの「山」において、一番高い $un(t_i, m_j, w)$ をとる単語を、その「山」における特徴的な単語として表示する。しかし、同一の単語が何度も同じ層上に表示されるのは不適切であり、多様であることが望ましい。そこで、小説冒頭に近い「山」から順に、表示する単語を決定し、すでに表示された単語の興

味喚起度 $at(t_i)$ を 0.1 倍し、同じ単語が複数回表示されないようにする。なお、その都度 $\max_{t \in T}(at(t), w)$ の値も更新している。

提案インタフェースにおいて、一つの層に表示させる単語の数は、最大 30 単語とし、高い層から優先的に表示させた。

5. 提案インタフェースの評価

提案インタフェースの有効性を示すために、被験者実験に基づく評価を行った。実験に用いる小説のデータは、青空文庫^(注2)で公開されている小説の内の、40 冊分を使用した。レビューは web 本棚サービスである Booklog^(注3)に投稿されているものを使用した。

読みたい小説かの判断には、文章の読みやすさ、物語の筋に対する興味等が影響すると考えられる。そしてその判断は、論理的に行われるよりも、むしろ、感覚的に行われると考えられる。我々は、各観点から、より確信を持った判断を下せるインタフェースは、より有効性が高いと考えた。そこで、本実験では、読みたい小説かどうかを、より確信をもって選択できたかでインタフェースの評価を行う。評価のためのベースラインとして、青空文庫の Web サイト上での書籍閲覧を利用する。

被験者には検索、推薦などにより、候補として提示された小説を、実際に読むかを判断するために「立ち読み」を行う状況を想定してもらった。それぞれの被験者に対して、金色夜叉(尾崎紅葉著)、フランケンシュタイン(メアリー・シェリー著)、ジーギル博士とハイド氏の怪事件(ロバート・ルイス・ステイヴンソン著)、真珠夫人(菊池寛著)の 4 冊のうち 2 冊を提案インタフェースで「立ち読み」してもらい、残りの 2 冊を青空文庫のサイトでブラウジングしてもらった。青空文庫では、小説本文のみが表示される。提案インタフェースで「立ち読み」する小説と、「立ち読み」を行う順序は我々が指定した。実験前に、被験者には提案インタフェースの操作に慣れてもらい、青空文庫を「立ち読み」する際には、冒頭から順に読む必要はないこと、本文検索機能を用いてよいことを説明した。「立ち読み」時間は一分間とし、各「立ち読み」終了後にアンケートに回答してもらった。アンケート内容は、小説家の代表作を評価した Novels and Novelist: A Guide to the World Fiction (Martin Seymour-Smith 編集) の評価軸(“読みやすさ”, “人物の魅力”, “プロット”, “文学的価値”)を参考にして以下の問を作成した。

Q1. 読みたい小説かどうかを判断できましたか？

Q2. “小説の読みやすさ(読みやすい文章かどうか)”を判断できましたか？

Q3. “小説の話の筋(テーマ・主題)”が魅力的かどうかを判断できましたか？

Q4. “小説の登場人物が魅力的かどうか”を判断できましたか？

Q5. “文学的価値があるかどうか”を判断できましたか？文学的価値の定義は、貴方が思う定義をお願いします。

表 1 「立ち読み」の小説とインタフェースの組み合わせ (人)

書籍タイトル	提案インタフェース	青空文庫	計
金色夜叉	4	3	7
真珠夫人	4	3	7
ジーギル博士と ハイド氏の怪事件	3	4	7
フランケンシュタイン	3	4	7
合計	14	14	28

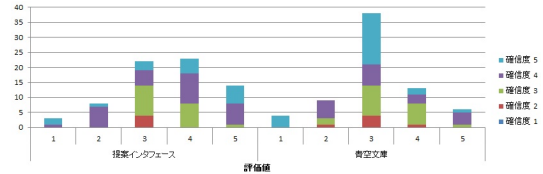


図 5 各手法における評価値毎の確信度の回答数

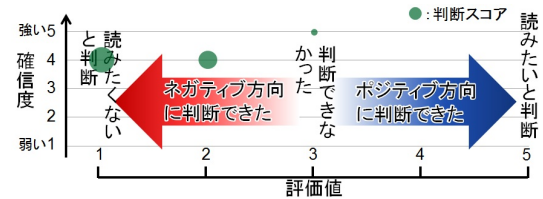


図 6 判断スコア

被験者には、それぞれの問に対して 5 段階(例: 5:読みたいと判断, 4:どちらかと言えば読みたいと判断, 3:判断できなかった, 2:どちらかと言えば読みたくなないと判断, 1: 読みたくなないと判断)で回答してもらった。また、回答をどれだけ確信を持てたかについて 5 段階(5:強い確信がある, 4:確信がある, 3:どちらでもない, 2:確信がない, 1: 全く確信がない)で評価してもらった。

6. 実験結果

被験者は 7 名が 4 冊の「立ち読み」を行った。小説とインタフェースの組み合わせに対して、「立ち読み」を行った被験者の人数を表 1 に示す。また、評価値毎の確信度をグラフにまとめたものを図 5 に示す。

6.1 選別の有効性の評価

確信を持った選別ができたかを、評価するために、被験者の回答から判断スコアを求める。被験者集合を P とした際に、設問 q の判断スコア (q) を以下の式で定義する。

$$\text{判断スコア}(q) = \sum_{p \in P} |3 - p_{\text{評価値}}| \times p_{\text{確信度}} \quad (4)$$

図 6 に示す通り、評価値は 3 を中心とし、5 がポジティブに評価(読みたいと判断, 魅力があると判断), 1 がネガティブに評価(読みたくなないと判断, 魅力がないと判断)となっている。同程度の確信度であっても、評価値が 3 から離れるほど、より強い確信を持って答えられたととらえることができる。また、確信度が 5 であっても、評価値が 3 である場合は、「判断できなかったことを、確信をもって答えた」ことを意味し、確信を持った選別を行えたとは考えにくく、判断スコアでは 0 となる。

(注2) : <http://www.aozora.gr.jp/>

(注3) : <http://booklog.jp/>

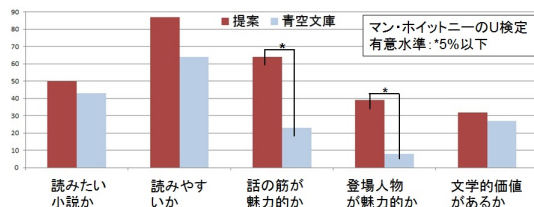


図 7 各設問の判断スコア

各設問の、判断スコアを図 7 に示す。提案インタフェースを用いることで、“小説の読みやすさ”、“小説の話の筋が魅力的か”、“小説の登場人物が魅力的か”についてより確信を持った判断が可能であることがわかった。インタフェースに対する被験者毎の回答、 $|3 - p\text{評価値}| \times p\text{確信度}$ 、の集合が正規分布に従わないことから、検定には、マン・ホイットニーの U 検定を用いた。検定により、“小説の登場人物が魅力的か”、“小説の話の筋が魅力的か”には有意水準 5 % で有意差があることが示された。

各質問の確信度を詳しく見るために、図 8 に、各質問の評価値毎の確信度を示す。

- “読みたい小説か”においては、提案インタフェースの方が、読みたい小説だと自信を持って判断できたという回答が多いが、大きな違いは見られなかった。
- “読みやすさ”においては、小説本文を読む時間が長い青空文庫の方が、より高い評価値と確信度をとっていたが、実際は異なっていた。その理由として、興味喚起度 MAP 中のキーワードが、文章内容の理解を助けたことが考えられる。
- “話の筋が魅力的か”においては、ユーザが提案インタフェースを用いることで、より確信をもって、話の筋が魅力的だと判断したことがわかった。興味喚起度 MAP において、小説全体の興味の推移と、各ピークでのキーワードを把握することで、読む文章量が青空文庫を用いた時よりも少なくとも、話の筋をより理解できたためだと考えられる。
- “登場人物が魅力的か”においては、提案インタフェースにおいて、より確信を持った判断ができた。また、提案インタフェースにおいて、より魅力的だと感じている理由としては、興味をひくと示された文章中の登場人物を読んだことが考えられる。
- “文学的価値があるか”においては、両インタフェースの評価値毎の確信度が似たような割合となった。本実験で用いた小説は、有名作品ばかりだが、それでも、文学的価値を判断することは、難しかったと考えられる。文学的価値を判断するためには、異なる支援が必要だが、「立ち読み」によって、文学的価値を判断することは、稀だと考えられるので、優先度の低い課題とする。

提案インタフェースにおいて、“話の筋の魅力”、“登場人物の魅力”について、魅力的、もしくは魅力的でないと、より確信を持って判断できることがわかった。しかしながら、最終的な意思決定にあたる“読みたい小説か”の判断スコアにおいては、インタフェース間に有意差が見られなかった。“読みたい小説か”と、他の 4 つの設問との間に、相関は見られなかった。”

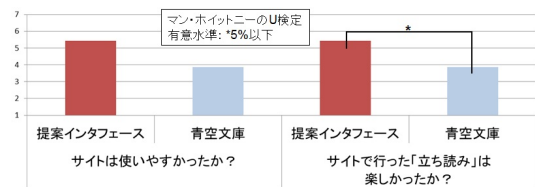


図 9 インタフェースに対するアンケート項目の回答の平均値

読みたい小説か”の選別の要因を明らかにすること、また話の魅力等が小説の選別に役立たせる手法を模索することを今後の課題とする。

6.2 提案インタフェースのユーザエクスペリエンスの評価

4 冊全ての「立ち読み」が終了した後に、「立ち読み」の感想に関するアンケートを行った。被験者には、青空文庫、提案インタフェースについて、以下の二つの間に 7 段階 (例：7:とても使いやすい, 6:使いやすい, 5:どちらかと言えば使いやすい, 4:どちらでもない, 3:どちらかと言えば使いにくい, 2:使いにくい, 1:とても使いにくい) で答えてもらった。

Q1. サイトは使いやすかったですか？

Q2. サイトでおこなった立ち読みは楽しかったかですか？

アンケート結果を図 9 に示す。両方の質問において、提案インタフェースの平均値がより高かった。また、マン・ホイットニーの U 検定において、“立ち読みの楽しさ”には有意水準 5 % で有意差があることが示された。

サイトの使いやすさにおいては、提案インタフェースは、操作できることが多いことと、慣れの点から青空文庫よりも低い値をとっていたが、結果はより高い値となった。その理由として、自由記述回答において、興味喚起度 MAP が「立ち読み」を開始する箇所の目安となったと述べられていることから、被験者は、操作性以外の利便性の要素も含めて、使いやすさを評価したことが考えられる。

「立ち読み」の楽しさにおいても、提案インタフェースはより高い値をとった。自由記述回答では、文章を難しく感じる小説に対して使ってみたいという意見があり、興味喚起度 MAP による支援が楽しさにつながったと考えられる。

7. ま と め

本論文では、電子書籍小説を対象とした「立ち読み」インタフェースを提案した。「立ち読み」インタフェースは、小説の本文と、レビューから自動的に生成される、興味喚起度 MAP を基盤としている。被験者実験により、確信をもった選別を行えたか、という観点から、インタフェースを評価した結果、提案インタフェースでは、「話の筋・登場人物の魅力」に対して、魅力的だと確信を持って判断できていたことがわかった。また、提案インタフェースにおいて、楽しく「立ち読み」ができることも示すことができた。提案インタフェースによって、Web においては「立ち読み」を行う人が少ない、また、選別に役に立つページを発見することが困難であるという問題点を、少なからず解決することができた。

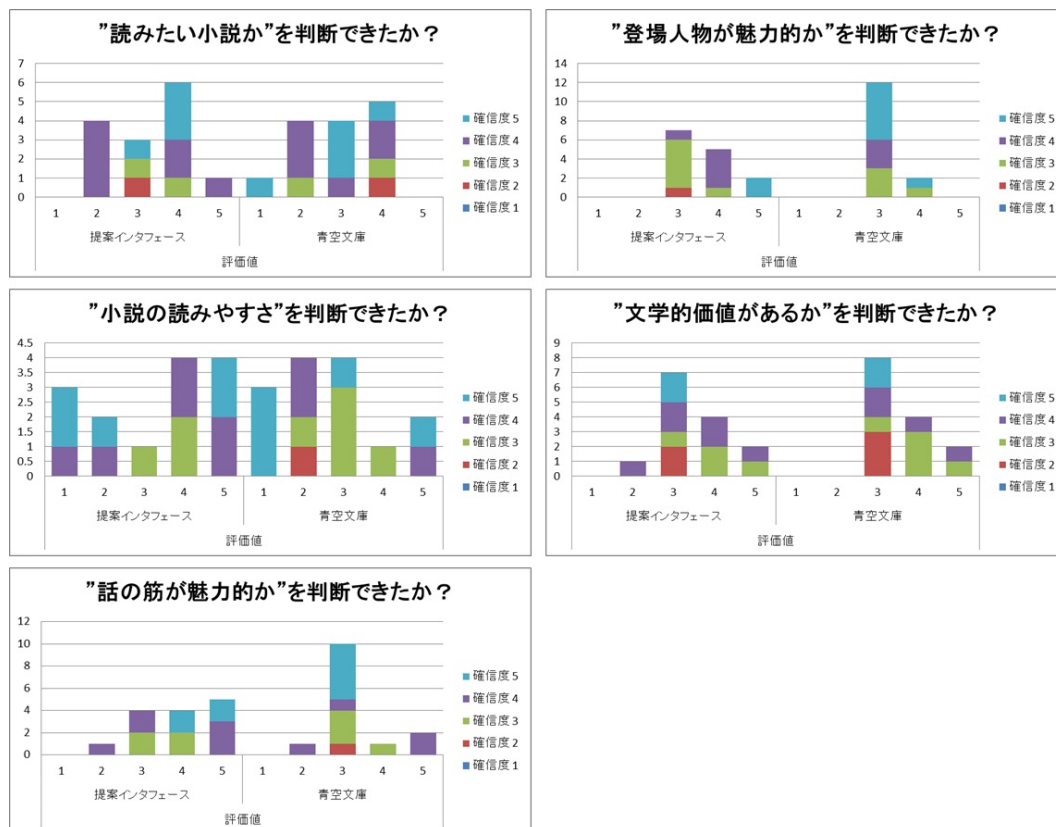


図 8 各手法における確信度と評価値の関係

文 献

- [1] 日本出版インフラセンター：電子出版と紙の出版物のシナジーによる書店活性化事業 (online),
- [2] 村上 憲郎：Google の切り開く情報の世界 プロフェッショナルの仕事とは, 情報管理, Vol 53, No 12, pp. 651–664 (2011).
- [3] 砂山 渡, 谷内田正彦：情報探索におけるブラウジング行動：文章の特徴を表すキーワードを発見して重要文を抽出する展望台システム, 電子情報通信学会論文誌. D-I, 情報・システム, I-情報処理, Vol 84, No 2, pp. 146–154 (2001)
- [4] 山本圭介, 棚瀬達史, 石戸谷顕太郎, 大平茂輝, 長尾確：映像と論文のアノテーションに基づく論文読解支援, 第 73 回全国大会講演論文集, Vol 2011, No 1, pp. 757–759 (2011)
- [5] 松田 千春：情報探索におけるブラウジング行動：図書館と書店における行動観察を基にして, 三田図書館・情報学会, No 49, pp. 1–31 (2003).
- [6] Hinze, Annika., McKay, Dana., Vanderschantz, Nicholas., Timpany, Claire and Cunningham, Sally Jo.: Book selection behavior in the physical library: implications for ebook collections, *JCDL '12*, pp. 305–314 (2012).
- [7] Ooi, Kamy.: How adult fiction readers select fiction books in public libraries: a study of information-seeking in context, Masters, Victoria University of Wellington, New Zealand(2008)
- [8] McKay, Dana.: A jump to the left (and then a step to the right): reading practices within academic ebooks, *OzCHI '11*, pp. 202–210 (2011).
- [9] 沢井 康孝, 山本 和英：文書に対する大衆の興味の強さの推定提案人工知能学会全国大会論文集 (CD-ROM) (2007)
- [10] 谷口 敏夫：『天人五衰』の小説構造可視化：三島由紀夫『豊饒の海』の絵解き, 京都光華女子大学研究紀要, Vol. 42, pp. A35–A57 (2004).
- [11] CPS：山田 美幸：プラネット・シェイクスピア：文学作品におけるストーリーの可視化 (online), (2011.02.23).
- [12] Plaisant, Catherine and Rose, James and Yu, Bei and Auvil, Loretta and Kirschenbaum, Matthew G. and Smith, Martha Nell and Clement, Tanya and Lord, Greg.: Exploring erotics in Emily Dickinson's correspondence with text mining and visual interfaces, *Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries.*, pp. 141–150 (2006).
- [13] Lohmann, Steffen and Burch, Michael and Schmauder, Hansjörg and Weiskopf, Daniel.: Visual analysis of microblog content using time-varying co-occurrence highlighting in tag clouds, *Proceedings of the International Working Conference on Advanced Visual Interfaces.*, pp. 753–756 (2012).
- [14] Lee, Bongshin and Riche, Nathalie Henry and Karlson, Amy K. and Carpendale, Sheelash.: SparkClouds: Visualizing Trends in Tag Clouds, *IEEE Transactions on Visualization and Computer Graphics.*, Vol 16, No 6, pp. 1182–1189 (2012).
- [15] Havre, Susan and Hetzler, Elizabeth and Whitney, Paul and Nowell, Lucy.: ThemeRiver: Visualizing Thematic Changes in Large Document Collections, *IEEE Transactions on Visualization and Computer Graphics.*, Vol 8, No 1, pp. 9–20 (2002).
- [16] Liu, Shixia and Zhou, Michelle X. and Pan, Shimei and Song, Yangqiu and Qian, Weihong and Cai, Weijia and Lian, Xiaoxiao.: TIARA: Interactive, Topic-Based Visual Text Summarization and Analysis, *ACM Trans. Intell. Syst. Technol.*, Vol 10, No 2, pp. 1303–1312 (2012).
- [17] 村井 聡一提案牛尼 剛聡：電子書籍小説の効率的な選別のための興味喚起度に基づく「立ち読み」インタフェース提案 WebDB Forum (2011).
- [18] 黒橋 禎夫, 白木 伸征, 長尾 真：出現密度分布を用いた語の重要説明箇所の特定, 情報処理学会論文誌, Vol 38, No 4, pp. 845–854 (1997).