

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 6 日現在

機関番号：17102

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24500119

研究課題名(和文)電子書籍小説の効率的な選別のためのレビューを利用した「立ち読み」インタフェース

研究課題名(英文)A browsing interface for effective selection of electronic novels based on user reviews

研究代表者

牛尼 剛聡 (Ushiana, Taketoshi)

九州大学・芸術工学研究科(研究院)・准教授

研究者番号：50315157

交付決定額(研究期間全体)：(直接経費) 4,200,000円

研究成果の概要(和文)：本研究では、Web上で閲覧できる電子化された小説を対象とし、「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をした。本インタフェースでは、小説の選別においては、小説の内容に興味をひかれるかどうかを重要と考え、「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マップを利用する。本インタフェースでは、ユーザが興味喚起度マップ上の単語をクリックすることによって、効率的に「立ち読み」を行うことができる。被験者による実験によって、本インタフェースを利用した「立ち読み」によって、Webブラウザを用いた「立ち読み」よりも、電子化された小説を効果的に選別できることが示された。

研究成果の概要(英文)：In this paper, we proposed a browsing interface for supporting a user to decide whether the user reads an electronic novel effectively on the web. Our interface provides an attractiveness map, which visualize a transition of attractiveness in a target novel, for effective selection of a novel. An attractiveness map has a layered structure and a section where high attractiveness words appear in high density has high value. The attractiveness of a term in a novel is estimated automatically from online book reviews of the novel. A user can browse the electronic novel effectively by clicking on attractiveness terms on the attractiveness map. Experimental results show that browsing of our interface supports users to decide whether the user would like to read a target novel or not more efficiently than browsing of a web browser.

研究分野：情報学

キーワード：インタフェース 電子書籍 立ち読み 選別 レビュー 可視化

1. 研究開始当初の背景

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、読みたい書籍を見つけることが困難になったため、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、ユーザが読みたい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。ユーザは、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、候補を絞る作業を行う。

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だと、即座に断定するわけではない。提示された候補の書籍に対して、何らかの方法で価値を判断し、選別を行う。我々は、効果的な選別方法の一つに、実際に書籍の本文を読む「立ち読み」があると考えた。

実世界での立ち読みと比較し、Web 上で「立ち読み」が行われない理由として、閲覧ページの制限が考えられる。実世界での立ち読みでは、好きなページを読めるのに対して、Web 上で提供されている「立ち読み」では、ユーザが閲覧可能な範囲が先頭ページから、多くとも数十ページに限定されていることが多い。そのため、ユーザの選別に役に立つ内容が書籍に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。

2. 研究の目的

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』と定義する。また、『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。

本研究では、電子書籍小説の「立ち読み」においては、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いう考えに基づいたインタフェースを開発する。

「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、一部分を読むことで、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 上にある読者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。

本論文で提案するインタフェースでは、興

味喚起度マップにより、ユーザが小説の全体的な流れを俯瞰し、興味を持った箇所を閲覧可能とする。

3. 研究の方法

本研究では、まずユーザレビューに基づいて興味喚起度マップを自動生成する手法を開発した。次に、興味喚起度マップ上で動作する電子書籍の立ち読みインタフェースを設計し、Web 上で動作するプロトタイプシステムを実装した。そして、実装したプロトタイプシステムを利用して、提案手法の有効性を評価する評価実験を行った。

4. 研究成果

4.1 興味喚起度マップ生成手法

興味喚起度マップを生成するために、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推定する。ここでは、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、小説の流れに従った興味喚起度の推移を可視化する。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、小説の流れに沿った興味喚起度の推移だけでは、そこにどのような内容が書かれているのかが解らない。そこで、最後に、それぞれの層のピーク箇所に特徴的なキーワードを決定し、配置する。

4.1.1 単語の興味喚起度の推定

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3種類のカテゴリに分類する。

1. 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)
2. 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)
3. 小説本文には出現しないが、そのレビューには出現する単語(カテゴリ C)

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。また、カテゴリ C に含まれる単語には、作者・小説に関する情報や、レビューの小説に対する意見や感情が含まれると考えられる。

我々は、「レビューが興味を感じた対象は、他の読者の興味をひく可能性が高い」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その小説における興味喚起度は 0 とする。カテゴリ C の単語は、小説に対する読者の印象を推定するために重要な情報が含まれていると考えられる。しかし、本研究において生成しようとしている興味喚起度マップは、小説の特定の部分にアクセスするための機構であり、ここでは、小説本文中に出現する単語の中で、ユーザの興味を喚起する単語を抽出することが目的であるため、カテゴリ C に含まれる単語は、興味喚起度推定の対象外とした。

本手法では、単語の興味喚起度を、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、読者の興味をひ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。この手法は、文書中の単語の重み付けのために一般的に利用されている TF-IDF 手法の考え方を発展させたものであり、小説とそのレビュー集合を利用して、小説に出現する単語の重みを計算するため手法であると位置付けられる。一般的な TF-IDF 手法では、同種の文書から構成される文書集合を使用するのに対して、上記の手法では小説の単語の重要度（興味喚起度）を推定するために、小説本文における出現頻度ではなく、レビュー集合における単語の出現頻度を利用する点が特徴である。

4.1.2 小説の特定箇所の興味喚起度とその変化

興味喚起度マップでは、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。我々は、小説を構成する最小の意味的なまとまりは、小説中の単語であると考え、「興味を喚起する単語を含む文章は、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を、小説として興味喚起する度合いが高い箇所と考える。形式的には、小説に出現する単語の出現位置を先頭から順番に i で表すとき、単語出現位置 i を中心とした部分の興味喚起度を、 $i-w$ から $i+w$ 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。ここで、 w は、一つの単語が影響を与える範囲を表し、文脈範囲と呼ぶ。本手法では、単語出現位置 i から離れた単語の重要度を軽く扱うためにハニング窓関数を各単語の興味喚起度に乗じた。

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。ハニング窓の幅が大き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。

4.1.3 ユーザに提示する特徴的な単語の決定と配置

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、特定の箇所にどのような内容の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。「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。

各「山」において、一番高い興味喚起度をとる単語を、その「山」における特徴的な単語として表示する。しかし、同一単語が何度も同じ層上に表示されてしまうと、各箇所の特徴の違いが見えなくなり、小説の全体的な流れの把握が困難になる。したがって、同じ層状に出現する単語にはある程度の多様性があることが望ましい。特徴的でない単語が出現する可能性もあるが、ユーザは様々な観点から興味を持つと考えられるので同じ単語が何度も同じ層状に表示されるよりも、ユーザが小説に興味を持つのに適していると考えられる。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を 0.1 倍し、同じ単語が複数回表示されないようにする。

4.2 実験と評価

評価は、単語の興味喚起度推定手法に関するものと、提案インタフェースに関するものの 2 種類の事項に関して行った。実験に用いる小説のデータは、青空文庫で公開されている小説の内、40 冊分を使用した。レビューは Web 本棚サービスである Booklog に投稿されているものを使用した。形態素解析には、オープンソースの形態素解析エンジン MeCab を用いた。

4.2.1 単語の興味喚起度推定手法に関する評価

本論文で提案した興味喚起度推定手法の有効性を評価するための実験を行った。評価のためのベースラインとして、小説本文を文書単位とした TF-IDF 法を用いた。

実験方法は以下の通りである。まず、被験者には、対象とした 40 編の小説の中から、任意の小説を選択してもらい、選択した小説において、提案手法による興味喚起度に基づく同一順位十位内の単語と、比較対象である

TF-IDF 値における同一順位十位内の単語を提示した。そして、それぞれの単語に対して、「その単語が出現する小説の箇所を讀んでみたいと思うか？」について 5 段階（「讀んでみたい」、「どちらかと言えば讀んでみたい」、「わからない」、「どちらかと言えば讀みたくないとは思わない」、「讀みたくないとは思わない」）で回答してもらった。なお、提示する単語の順番はランダムに決定し、両手法において同じ単語が含まれている場合は、その単語を一度のみ提示した。

被験者 15 人が評価した小説の総数は 174（うち未読小説に対する評価:126）であった。提示した単語に関する評価「讀んでみたい」から「讀みたくないとは思わない」の選択肢に対して、それぞれ 5 点から 1 点の評価値を割り当て、各手法の書籍毎の平均評価値を計算した。

実験結果から、今回対象とした全ての場合において、提案する興味喚起度に基づいて選択された単語は、TF-IDF 法に基づいて選択された単語よりも、被験者の興味を喚起する程度が大きかった。そして、それらの結果には有意水準 1% で有意差があることが示された。したがって、提案する単語に対する興味喚起度推定手法は有効であることが示された。

次に、登場人物名が未読者に対して興味を喚起するかについて考察する。これは、登場人物名は、高い興味喚起度と TF-IDF 値を持つことが多いが、未読の書籍に対しては、ユーザの興味を喚起する可能性が少ないと予想されるためである。

分類した結果より、既読小説に関しては、登場人物名とそれ以外の単語の評価値はほとんど変わらなかった。しかし、未読小説に関しては、明らかに登場人物名に関する評価値はそれ以外の単語の評価値よりも低いことがわかった。したがって、未読小説に関しては、人名はユーザの興味を喚起しない可能性が高いと考えられる。なお、未読者においても、「フランケンシュタイン」、「ごん狐」、「三四郎」といった、書籍名に含まれたり、一般的に認知度が高い人名に関しては評価値が 4.0 点以上あり高かった。一方「小川」、「鈴木」、「よし子」といった、一般的な名前には、評価値が 1.0 と低かった。

4.2.2 提案インタフェースの評価

讀みたい小説かを判断するためには、文章の読みやすさや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、直感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、讀みたい小説かどうかを、より確信をもって選択できたかどうかという観点から提案インタフェースを評価する。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。青空文庫の Web サイト上での書籍閲覧をベースラインとした理由は、そこでは小説本文のみが表示されており、電子化された小説の閲覧環境として最も一般的であると考えられるからである。

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)、ジーキ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。提案インタフェースで「立ち読み」する 2 タイトルと、青空文庫で「立ち読み」をする 2 タイトルの組み合わせ方、および 4 冊の「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 Novelists: A Guide to the World Fiction の評価軸(読みやすさ、人物の魅力、プロット、文学的価値)を参考に以下の間を作成した。

[Q1.] 讀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、貴方が思う定義でお願いします。

被験者には、それぞれの間に対して、5 段階の評価値(例: 5: 讀みたいと判断, 4: どちらかと言えば讀みたいと判断, 3: 判断できなかった, 2: どちらかと言えば讀みたくないと判断, 1: 讀みたくないと判断)で回答してもらった。また、回答をどれだけ確信を持てできたかについて、5 段階の確信度(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)で回答してもらった。

被験者 11 名が 4 冊の「立ち読み」を行った。

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 $ds(q)$ を定義した。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。

提案インタフェースを用いることで, 「読みたい小説かどうか」, 「小説の話の筋が魅力的かどうか」, 「小説の登場人物が魅力的かどうか」について, 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, 及び判断スコアの分布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, これらには有意水準 1% で有意差があることが示された。

4.2.3 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 段階で回答してもらった。

[Q1.] 使いやすかったですか?

[Q2.] 立ち読みは楽しかったかですか?

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, 「サイトの使いやすさ」には有意水準 5% で, 「立ち読みの楽しさ」には有意水準 1% 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, 青空文庫のインタフェースよりも, 実行できる機能が多いことや, ユーザが操作に慣れていないことから, 青空文庫よりも低い値をとると予想していたが, 結果は高い値となった。その理由として, 自由記述回答において, 興味喚起度マップが「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, 青空文庫のインタフェースよりも高い評価を得た。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ようという意見があり, 興味喚起度マップによる支援が楽しさにつながったと考えられる。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 6 件)

1. Taketoshi Ushiana and Noboru Matsuo. 2015. A method for finding similes to support final selection from search results. In Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication (IMCOM '15). ACM, New York, NY, USA, , Article 59 , 4 pages. DOI=10.1145/2701126.2701197 <http://doi.acm.org/10.1145/2701126.2701197>
2. 村井聡一, 牛尼剛聡: 電子化された小説の選別を支援する「立ち読み」インタフェース, 情報処理学会論文誌: データベース, Vol. 6, No. 4, pp. 24-34, (2013)
3. Soichi Murai, Taketoshi Ushiana: Review-based recommendation of attractive sentences in a novel for effective browsing, International Journal of Knowledge and Web Intelligence, Vol. 3, No. 1, pp. 58 - 69 (2012)
4. Souichi Murai, Taketoshi Ushiana: An Interface for Browsing Electronic Novels Using Attractiveness Map, Proc. of 16th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2012), pp. 1061-1070 (2012). DOI:10.3233/978-1-61499-105-2-1061
5. Souich Murai and Taketoshi Ushiana: A Method for Extracting Attractive Sentences from an Electronic Book Based on Reviews for Effective Browsing, Proc. of 15th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2011), pp. 24-33, Kaiserslautern, Germany (2011). DOI: 10.1007/978-3-642-23866-6_3
6. 村井 聡一, 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", WebDB Forum 2011 論文集, CD-ROM, 2011, 11 月

〔学会発表〕(計 8 件)

1. Taketoshi Ushiana and Noboru Matsuo. 2015. A method for finding similes to support final selection from search results. In Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication (IMCOM '15).
2. Souichi Murai, Taketoshi Ushiana: An Interface for Browsing Electronic Novels Using Attractiveness Map, 16th

- International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2012)
3. Souich Murai and Taketoshi Ushima: A Method for Extracting Attractive Sentences from an Electronic Book Based on Reviews for Effective Browsing, 15th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2011)
 4. 村井 聡一, 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", WebDB Forum 2011
 5. 村井 聡一, 牛尼 剛聡: 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, 第5回データ工学と情報マネジメントに関するフォーラム, 2013年3月.(学生プレゼンテーション賞受賞)
 6. 村井 聡一, 牛尼 剛聡, ネタばれと興味喚起を考慮した小説レビューのランキング手法, 第4回データ工学と情報マネジメントに関するフォーラム, 2012年3月.
 7. 村井聡一, 牛尼剛聡: 電子書籍小説の選別における興味喚起度 MAP の有効性の評価, 第155回データベースシステム研究会報告, Vol. 2012-DBS-155, No. 2, 2012年11月12日.(学生奨励賞受賞)
 8. 村井聡一, 牛尼剛聡: レビューを利用した電子書籍立ち読み支援, 第3回データ工学と情報マネジメントに関するフォーラム, 2011年3月.

〔図書〕(計 0件)

〔産業財産権〕

出願状況(計 0件)

取得状況(計 0件)

〔その他〕

6. 研究組織

(1)研究代表者

牛尼 剛聡 (USHIAMA, Taketoshi)

九州大学・大学院芸術工学研究院・准教授

研究者番号:

(研究協力者)研究代表者

村井 聡一 (MURAI, Soichi)

松尾 昇 (MATSUO, Noboru)

趙 耕弘 (ZHAO, Genghong)