

小説における振り返り支援システムの提案

竹村徳文^{†1}河合由起子^{†1}熊本忠彦^{†2}白石優旗^{†3}張建偉^{†3}^{†1} 京都産業大学^{†2} 千葉工業大学^{†3} 筑波技術大学

1 はじめに

携帯端末の普及に伴い、電子書籍における多種多様な閲覧サービスが広がっている。電子書籍端末の Kindle ではページに対するブックマークが可能であり、それらをユーザ間で共有することもできる。こういった従来の閲覧サービスでは、検索機能が提供されており、特定のキーワードを含むページを容易に検索できるが、読書中に失念したような事柄を検索するのは、キーワードそのものを思い出せないため、容易でない。そこで本研究では、電子書籍のうち小説を対象とし、小説の内容理解およびクライマックスにおける感動を増幅させるための振り返り支援システムを提案するとともに、構築したシステムを用いてその有用性を検証する。

具体的には、閲覧中のページに登場する事柄のうち、重要かつ失念率の高い事柄を抽出し、提示するとともに、その事柄が登場した特徴的な既読ページを推薦することで、内容理解を支援する。ただし、提示する事柄によってはクライマックスにおける感動を減少させる要因となりうるため、クライマックスに重要な事柄を抽出し、その事柄の印象を逆に薄めることで、クライマックス時の感動を増幅させる。

2 関連研究

読書における内容理解の支援として、読書を中断した際に挟んだ「しおり」の位置から最初のページまでの既読部分のあらすじを生成し、提示するという研究がある [1]。また、クライマックスにおける感動増幅として、ネタバレ防止は有用である。スポーツの勝敗にまつわるネタバレ防止手法 [2] では、ネタバレ情報の動的フィルタリングおよびネタバレ情報に対する重要度を検証している。小説を対象としたジャンル推定と人物抽出 [3] では、人名を抽出しその周辺文脈から人物の特徴を表す表現を抽出する。

3 振り返り支援システム

提案する振り返り支援システムは、現在閲覧しているページ上に内容理解支援のための特徴語と感情増幅のための特徴語をいくつか提示する。ユーザは、特徴

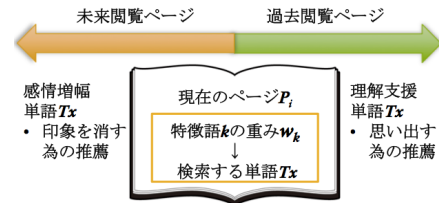


図1: 内容理解支援と感情増幅のための振り返り

語を選択することで、その特徴語に関連するページを振り返ることができる。なお、感情増幅のための特徴語は、クライマックス以降に提示される。

内容理解支援と感情増幅を実現するため、現在閲覧しているページを基準に既読（過去閲覧）ページおよび未読（未来閲覧）ページに出現する単語間のページ間隔を考慮する（図1）。既読ページの分析では、出現頻度が高いにもかかわらず、長く出現していない単語ほど読者の想起要求は高いと仮定し、閲覧ページに出現している単語のうち、既読ページ中に既に出現している単語までのページ間隔も考慮する。未読ページの分析では、真相提示等のクライマックスに深く関わる重要度の高い単語を抽出し、その単語をクライマックスまで推薦提示しないことで単語の印象を薄め、読者の感動を高めることを試みる。

3.1 内容理解支援のための振り返り

内容理解支援では、読者が閲覧中のページ p_i までに失念した単語を特徴語として抽出する。まず、 p_i における単語 $word_k$ の重要度 w_k を以下の式より算出する。

$$w_k = \frac{word_k \text{ の出現回数}}{p_i \text{ 中の単語総数}} \cdot \log \frac{\text{ページ総数}}{word_k \text{ の出現ページ数}} \quad (1)$$

次に、既読ページ $p_1 \sim p_{i-1}$ において $word_k$ が最後に出現したページ p_j を抽出し、閲覧中のページ p_i との差分 $(p_i - p_j)$ を算出した後、この差分値をページ p_i までの総ページ数 n で割った値を重要度 w_k に掛け合わせ、 x を求める。この x が以下の式を満たすようであれば、読者が失念している事柄とし、内容理解支援のための特徴語とする。なお、 T_x は別途定めた閾値を表す。

$$T_x \geq \frac{1}{(1 + e^{-x})}, \quad x = w_k \cdot \frac{(p_i - p_j)}{n} \quad (2)$$

式に標準シグモイド関数を用いたのは、直近の既読ページや遠い過去のページの重みを小さく、中間ページの重みを大きくするためである。

3.2 感情増幅のための振り返り

感情増幅では、今後重要となる事柄を抽出するために、閲覧中のページ p_i までの総ページ数 n を単語 $word_k$ が

A Proposal of a Support System for User Look-back Behavior on a Novel

^{†1} Narufumi TAKEMURA ^{†1} Yukiko KAWAI ^{†2} Tadahiko KUMAMOTO ^{†3} Yuhki SHIRAISHI ^{†3} Jianwei ZHANG^{†1} Kyoto Sangyo University^{†2} Chiba Institute of Technology^{†3} Tsukuba University of Technology

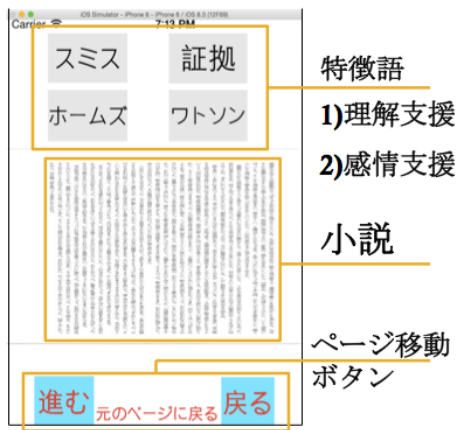


図2: 振り返り支援システムのインターフェース

次に出現する未読ページ p_j とページ p_i との差分 ($p_j - p_i$) で割った $x = w_k \cdot n / (p_j - p_i)$ の値を用いる。この値は小説のクライマックスであるかそうでないかで場合分けされ、 T_x 以上かつクライマックス以前である場合は特徴語として推薦提示しない。クライマックス以後であれば式 (2) より抽出した特徴語を提示する。

4 システムの実装と評価

今回提案した振り返り支援システムを Objective-C で構築し、抽出した特徴語を検証した。

4.1 インタフェース

図2に構築した振り返り支援システムの iPhone(iOS) 上のインターフェース提示例を示す。中央に小説のページ、下部にページ移動ボタンを配置した。上部には抽出した特徴語を4つまで提示する。

4.2 内容理解支援と感情増幅の検証

シャーロック・ホームズシリーズの推理小説¹から抽出された内容理解支援もしくは感情増幅のための特徴語を20代の男性4人、女性2人、計6名が評価した。なお、特徴語抽出の際の閾値 T_x は1.98とした。その結果を図3(縦軸は適合率)に示す。

抽出された内容理解支援のための特徴語は、ページあたり平均20語であり、感情増幅のための特徴語はページあたり平均18語であった。感情増幅のための特徴語は、真相提示ページでの伏線となるため、クライマックスページのみで検証を行った。なお、今回のクライマックスとなる真相提示のページは著者の主観に基づいて決定した。適合率は(正解と判定された単語の数) / (システムが推薦した単語の数)とし、正解判定は各被験者が行った。このとき、各ページから抽出された特徴語が「忘れたので理解したい単語」か「理解しているがもう一度確認したい単語」か「振り返る必要のない単語」かを評価してもらい、前の2つを正解とした。

¹ Arthur Conan Doyle 著「瀕死の探偵」

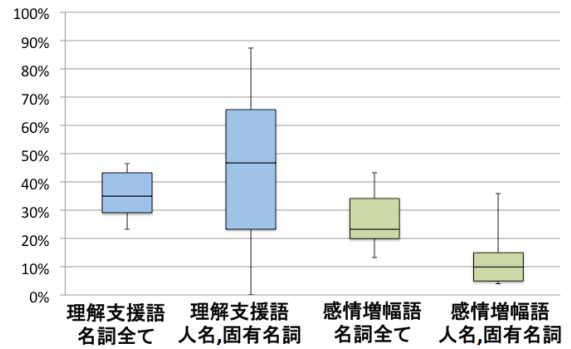


図3: 評価結果

内容理解支援評価では、特徴語の品詞を名詞とした場合の適合率の中央値は35.0%であったが、特徴語の品詞を名詞の中の人名と固有名詞に絞ることで、46.8%に上昇した。一方、感情増幅の評価では、特徴語の品詞を名詞とした場合の適合率の中央値は23.3%であったが、人名と固有名詞に絞ると、逆に10.0%に下がった。

今回の評価では十分に高い適合率が得られなかった。その原因として、短編推理小説を対象に評価を行ったため、登場人物が7人程度と少なく、犯人が小説の中盤から連続して登場していたことやクライマックスにおける凶器や犯行現場といった状況を考慮していなかった点が挙げられる。なお、状況の抽出手法としては、単語間の共起関係等の利用が考えられる。

5 おわりに

本研究では、小説(電子書籍)を対象に、失念した可能性の高い重要な単語を提示し、その単語が出現する特徴的なページの振り返りを支援するとともに、クライマックスで重要な単語をクライマックスまで推薦しないことで感動を増幅させるシステムを提案した。

また、評価実験より、内容理解支援に対しては適合率が50%程度であったが、感情増幅に対しては30%以下であることがわかった。

今後の課題として、感情増幅のための状況を表現する特徴語抽出による精度向上が挙げられる。

謝辞

本研究の一部は、JSPS 科研費 15K00162, 16H01722 の助成を受けたものである。ここに記して謝意を表す。

参考文献

- [1] 森晴菜, 山西良典, 西原陽子, 福本淳一, “しおりの前後での単語重要度を用いた小説の既読部分からのあらすじ生成”, ARG SIG W12, No.9, W12-2016-25, 2016.
- [2] 中村聡史, 小松孝徳, “スポーツの勝敗にまつわるネタバレ防止手法の検討”, 情報処理学会インタラクティブ2012.
- [3] 馬場こづえ, 藤井敦, 石川徹也, “小説テキストを対象としたジャンル推定と人物抽出”, 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 (DEIM 2012), F7-2, 2012.