

F-02

電子教科書におけるユーザの下線付加行動に基づく教材検索方式

Teaching material retrieval method based on underlining in e-book

大見 真一† 北山 大輔‡
Shinichi Ohmi Daisuke Kitayama

宮本 節子† 角谷 和俊‡
Setsuko Miyamoto Kazutoshi Sumiya

1. はじめに

近年、Apple の iPad や Amazon の kindle など、電子書籍端末が急激に普及している。また電子書籍端末を教育へ利用する動向が活発である。図 1 のように iPhone や iPad を使用して講義資料のパワーポイントと講義映像を同時に視聴しながら、パワーポイント資料に下線を引いたり、コメントを書き込むことができるモバイル環境の講義視聴システムが実験的に利用されている[1]。このようなシステムを用いることで、その書き込みなどを確認しフィードバックを行うなど、学生と教員のコミュニケーションを補助することが可能である。電子書籍端末を教育の場面で利用する背景に、以下の特徴が挙げられる。原文資料に対して下線や書き込みなどのアノテーション付与が直感的で容易である。また、付与したアノテーションデータを蓄積することが可能である。しかし、このようなアノテーション情報を蓄積することができる環境が整っているにも関わらず、アノテーション情報が学習の際に十分に活用されていない。本研究では、ユーザが付加したアノテーション情報を蓄積したデータベースを用いて、ユーザが下線を引いた単語ではなく、引いていない単語の重要度に着目し、下線を引いた操作情報に基づき個別のユーザに応じた学習教材検索へ応用する手法を提案する。



図 1 iPad による学習支援システム

2. 本研究概要とアプローチ

本章では、下線付加行動の学習方法の有効性について述べる。次に下線付加を利用した関連研究について述べる。

2.1 本研究概要

教科書などのテキストを用いて学習や読書をする際、ユーザは後で見直すために原文に下線を引く[2]。これらの下線を引く行動はテキストに対するユーザの情報探索・情報選択行動である[3][4][5][6][7][8]。本研究ではユーザが下線を引くという行動に基づき、各単語の重要度

を算出し、ユーザの学習進捗状況に合った教材の検索を行う。ユーザの下線付加行動に基づき単語の重要度を算出するために、以下の二つの尺度を定義する。

- i)隣接重要度
- ii)順序重要度

隣接重要度とは、下線が付加されていない語に対する重要度のことである。順序重要度とは、既に下線が付加された語に対する重要度である。例えば、日本史の教科書コンテンツにおける幕末の箇所「西郷」、「坂本」、「木戸」という 3 単語が近傍で記載されている場合を想定する。ユーザが「坂本」だけに下線を付加した場合、ユーザは「西郷」と「木戸」は下線が付加された「坂本」に比べ、ユーザにとって重要度が低かったため下線が付加されなかったと考えられる。さらに「木戸」に下線が追加された場合、ユーザにとって学習上、重要度が高い単語に新たに気付いたためであると考えられる。このように下線を引いた単語「坂本」から下線を引いていない単語「西郷」下線を追加した単語「木戸」を比較した場合、「西郷」、「木戸」、「坂本」の順に重要度が高いと推測することが可能である。このような二つの尺度を用いて、単語の重要度を算出する。この重要度に基づきユーザが学習すべき単語を推薦する。そして推薦された単語に基づき、教材へのリンク生成や問題作成を行うことで、電子教科書を用いた学習支援を行う。

2.2 関連研究

福永らは教師の下線と生徒の下線を比較し評価する方法を提案している[9]。本研究は学習者の下線情報だけを用いて評価している点で異なる。また鈴木らは下線の周辺の語句の意味からユーザの検索意図を推測し適切な検索クエリを生成し的確な絞り込み検索を行う方法を提案している[10]。本研究では下線の単語の意味を考慮せず、単語間の距離と下線を付加する順序でユーザにとって重要であると考えられる単語を推測している点で異なる。本研究ではユーザによって下線を付加する過程が異なる点に着目し、ユーザがテキストを読解する過程で対象の内容を理解している程度を推測し、ユーザの読解の状況に応じた最も適した学習教材の推薦を行う。

3. 下線重要度算出方式

本研究ではユーザが電子書籍端末を用いて、電子教科書内で重要と考える個所に下線を付加することを想定している。本手法ではユーザが下線を付加すると、下線が付加された語と付加されていなかった語に対して自動的に重要度を付与する。また、すでに下線が付加された単語が存在する箇所の単語に対し、新たに下線が付加される場合も考えられるため、隣接重要語と順序重要度の 2 種類の尺度を提案する。本章ではまず一つ目の尺度として、

†兵庫県立大学大学院環境人間学研究所、Graduate School of Human Science and Environment, University of Hyogo

‡兵庫県立大学環境人間学部、School of Human Science and Environment, University of Hyogo

対象範囲を初めて読書する際にユーザが下線を付加した単語とその近傍で下線を付加しなかった単語との距離に基づき算出される隣接重要度について述べる。次に、すでに下線を付加した単語が存在する対象範囲を再び読書する際に追加した下線に基づき、算出される順序重要度について述べる。

3.1 隣接重要度

隣接重要度は対象パラグラフを初めて読書して単語に下線を付加した際、下線が付加された単語の近傍に出現しているにも関わらず下線が付加されなかった単語に対して付与する重要度のことである。隣接重要度はユーザによって下線が付加された単語の近くに出現する語ほど低くなる。例えば、日本史の教科書コンテンツの幕末に関する記述の「西郷」「坂本」「木戸」という単語が近傍に記載されているパラグラフにおいて、ユーザが「坂本」という単語だけに下線を付加し、「西郷」と「木戸」には下線を付加しなかった。このような場合、ユーザは下線を付加した単語の前後に出現する単語が下線を付加した単語に比べて重要性が低いと判断したと考えられる。そこで下線が付加された単語「坂本」の近傍に出現する単語「西郷」「坂本」の隣接重要度は最も低くなる。隣接重要度 $neighbor_importance$ は以下の式により算出する。

$$dis(k_s, k_i) = |order(k_s) - order(k_i)| \quad (1)$$

$$neighbor_importance(k_s, k_i) = -\left(\frac{1}{1+\alpha}\right)^{dis(k_s, k_i)-1} \quad (2)$$

k_s は下線が付加された単語を表し、 k_i は下線が付加されていない単語を表す。 $order$ はパラグラフ内での単語の出現位置を表し、パラグラフの最初にある単語を 1 とする。 dis は下線付加単語から未下線付加単語までの距離であり、両者の出現位置の差を算出することで求める。 α は対象範囲内に出現する文字数の違いを考慮するための定数である。

図 2 は隣接重要度をグラフ化したものである。 k_1 から k_7 までの単語が近傍に記載されているパラグラフにおいて、 k_4 に下線を引いた場合、 k_4 の隣接に出現する k_3 、 k_5 の隣接重要度が -1 と最も低くなり、 k_3 から k_1 へ、 k_5 から k_7 へ k_4 から離れるに従い 0 に近くなる。

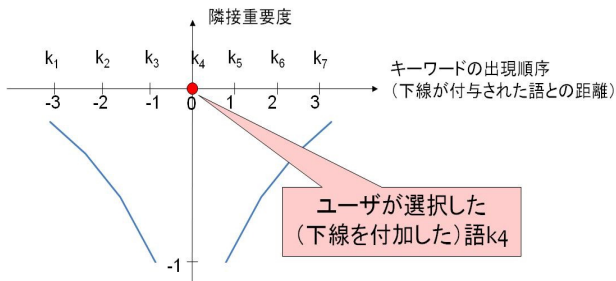


図 2 隣接重要度

3.2 順序重要度

順序重要度は既に下線を付加した単語が存在するパラグラフ内で、新たに下線が付加された単語に対して付与する重要度である。順序重要度は既に下線が付加されて

いる単語と異なる単語に下線が追加された場合、それまでに下線が付加された単語の重要度を低くする。ただし、既に下線が付加されている単語と同じ単語に下線が追加された場合、その単語に対する重要度は変更しない。例えば、日本史の教科書コンテンツの幕末に関する記述の「西郷」「坂本」「木戸」という単語が近傍に記載されているパラグラフにおいて、ユーザが初めての読書で「坂本」に下線を付加しており、読み返した際に新たに「木戸」に下線を付加したとすると、このユーザは「木戸」の重要性に気づき下線を追加したと考えられる。この場合、ユーザにとって以前に下線が付加されていた「坂本」よりも、新たに下線が付加された「木戸」に対する重要性が高くなったと考えられたため、「坂本」に対する順序重要度は下がる。さらに、「坂本」と「木戸」に下線が付加された状態で引き続き、ユーザが新たに「坂本」に下線を付加した場合、ユーザは下線を追加した「坂本」はそのパラグラフにおいて重要であると気づき、以前と同じ単語に対して下線を付加したにも関わらず、再度確認したと考えられる。この場合、ユーザにとって、初めに下線を付加した「坂本」に対する重要性は変わらないと考えられるため、順序重要度は変更しない。ただし「木戸」に対する重要度は下がる。順序重要度 $order_importance$ は以下の式により算出する。

$$order_importance(k_i^j, k_a^b) = \begin{cases} 0 & (\text{when } j = b) \\ importance(k_i^j) - \beta & (\text{when } i \neq a) \\ importance(k_i^j) & (\text{when } i = a) \end{cases} \quad (3)$$

k_a^b は既に下線が引かれている算出対象の単語を表し、 k_i^j は新たに下線が追加された単語を表す。 i および a はキーワードの種類を表す添字であり、同じ単語のキーワードは同じ数字が与えられる。また、 j および b はキーワードの ID であり、すべてのキーワードに異なる数字が与えられる。 β は定数である。 $importance$ はキーワードがすでに持っているに付与されている重要度を返す関数である。キーワード ID が同じである時、すなわち現在下線を付加した単語が算出対象である場合には順序重要度は 0 になる。算出対象のキーワードが下線を引いたキーワードと異なる単語である場合、重要度を β により減算する。算出対象のキーワードが下線が付加されたキーワードと同じ単語である場合、重要度は変化しない。

4. 下線重要度の例

本手法がユーザ意図を反映した単語重要度算出手法として有効であるか確認を行うために、従来手法として下線が付加されている単語に 1 点、付加されていない単語に 0 点を与える手法を用いて単語に重要度を付与した結果との比較を行った。

4.1 隣接重要度の考察

従来手法では未下線付加単語の重要度算出は行われないため、未下線付加単語を学習の対象として効率的に推薦することができない。しかし本提案手法では未下線付加単語に対して隣接重要度として重要度を付加することが可能である。

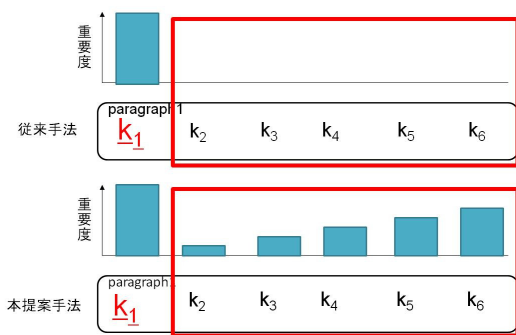


図3 隣接重要度算出の例

図3のように、 k_1 に下線を付加した場合、従来手法では k_2 から k_6 には重要度が付与されず、学習の対象にすべき単語を判定することができない。しかし提案手法では k_2 は k_1 へ下線を付加した際に学習されていると考えられるため、最も低い隣接重要度を付与する。

4.2 順序重要度の考察

従来手法では下線が追加された単語であっても重要度に差がないため、ユーザの学習進捗状況が考慮されていない。一方、提案手法では下線の付加順序が新しい単語ほど学習上確認の必要性が高い単語であり、下線の付加順序が古い単語ほど学習は完了していると考えられるため重要度を低くする。

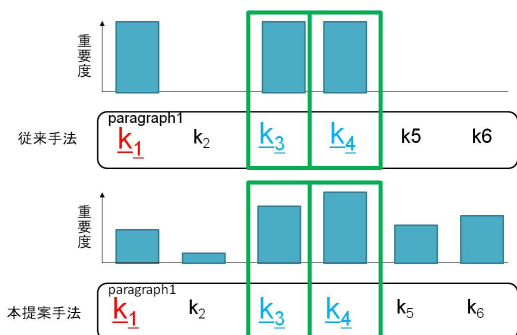


図4 順序重要度の算出例

図4では、既に k_1 に下線が付加されており、 k_3 に続いて k_4 に下線が追加されたことを表している。 k_4 が下線の付加行為が新しいということは、このユーザは k_4 が重要であると気付いたのは学習進捗上新しいということがいえる。つまり、このユーザは k_4 の確認をする必要性が高く、次に k_3 の重要度が高いといえる。このように本提案手法ではユーザの学習進捗状況に合わせて、単語の重要度を変化させることができる。従来手法では k_3 と k_4 の重要度に差はできない。

4.3 隣接重要度と順序重要度の考察

従来手法では下線付加単語は重要度が高く、未下線付加単語の重要度は0で、未下線付加単語は学習の対象として推薦されることはなかった。しかし本提案手法では未下線付加単語であっても、下線付加単語よりも重要度が高くなる可能性がある。それは初読で下線が付加され、再読の下線が多く追加された場合、初読の下線付加単語とその初読の下線付加単語から一定距離が離れている未下線付加単語との間で起こる。下線付加単語の学習が既に

十分であり、未下線付加単語が学習上まったく対象となっておらず、確認する必要がある場合に未下線付加単語の重要度が下線付加単語の重要度を上回る。

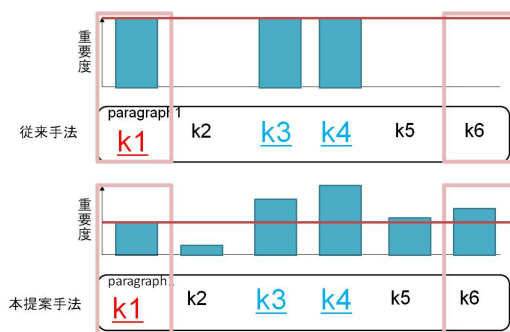


図5 隣接重要度と順序重要度の算出例

図5では、未下線付加単語の k_6 の重要度が下線付加単語の k_1 の重要度よりも高くなっている。本来ならば、下線付加単語の k_1 の方が重要になると考えられるが、この場合はそうではない。初読で k_1 に下線を付加してから、次に k_3 、 k_4 と下線を追加をしている。順序重要度により k_1 の重要度は低くなる。しかし初読で一定距離離れ高い隣接重要度が算出されており、 k_6 は重要度が高いままである。そこで下線追加が行われても、 k_6 は重要度が低くなる対象ではないので、高い重要度を保持したままである。つまりこのユーザは k_6 に対して、注意をしていないと考えられる。学習進捗上で、下線付加単語よりも未下線付加単語を学習する必要があるため重要度が高くなる。

5. まとめ

本手法ではユーザの下線付加行動に基づき単語の重要度を算出し、重要語を抽出する手法を提案した。具体的には隣接重要度を用いることで、ユーザが下線を付加していない単語の重要度を算出し、また順序重要度を用いることで、ユーザが下線を付加する以前に下線が付加されていた単語の重要度を算出する。算出された重要度に基づき単語をランキングすることができる。そのランキングは対象範囲のすべての単語である下線付加単語と未下線付加単語とすることができ下線付加単語よりも注意していない可能性のある未下線付加単語を効率的に推薦することができる。

今後の課題として次の内容を検討している。

- ・下線モデルの拡張
現在のシステムでは単語を下線の対象としているのみである。しかし文節、文章単位のような下線を付加するユーザも考えられるので、下線長への対応をモデル化する
- ・下線重要度を用いた教材検索方式の提案
単語の重要度に基づき抽出した重要語から検索質問を自動で生成するための手法を開発し、問題や参考図書へのリンクを生成してユーザの学習を支援する教材の検索を行い、電子書籍端末を利用した教育支援方法の確立を目指す。
- ・評価実験
本提案手法のプロトタイプを作成し、学習に有効性が認められるかアンケート実施する。高校生と大学生の2つ

の実験群を比較し、本提案手法が学習年齢により有意差が認められるか検討する。

6. 謝辞

本研究の一部は、平成 22 年度科研費基盤研究(B)(2)「ユーザの潜在的意図を用いたレス・コンシャス情報検索基盤の構築」(課題番号: 20300039)および平成 22 年度特別研究員奨励費(21.197)によるものです。ここに記して謝意を表すものとします。

7. 参考文献

[1]http://www.kyoto-u.ac.jp/ja/news_data/h/h1/news7/2010/100604_1.htm

[2]齋藤ひとみ・源田雅裕『ノートテイキングにおける方略使用の効果に関する検討』日本教育工学会論文誌 31(Suppl.), 197-200, 2008-02-10

[3]Annotation : From Paper Books to the Digital Library MARSHALL C.Proceedings of the 1997 ACM International Conference on Digital Libraries (DL 97), 1997

[4]魚崎 祐子, 伊藤 秀子, 野嶋 栄一郎『テキストへの下線ひき行為が内容把握に及ぼす影響』日本教育工学雑誌 26(4), 349-359, 2003-03-20

[5] Gynn, S. Capturing reader's attention by means of typographical cueing strategies, Educational Technology, 1978.

[6] Blanchard, J. and Mikkelsen, V. Underlining performance outcomes in expository text. Journal of Educational Research, 80: 197-201

[7] 松岡有希, 坂本竜基, 中田豊久, 伊藤禎宣, 武田英明: 論文概要に対する色付きアンダーライン付与システムの運用・分析, DEWS2006 論文集 (2006)

[8] プレゼンテーションコンテンツのためのコンテキストと視聴時行為に基づく区間検索方式 北山 大輔, 角谷 和俊 電子情報通信学会 第 19 回データ工学ワークショップ (DEWS2008)論文集, E4-6, 2008 年 3 月

[9]福永 良浩, 平嶋 宗, 竹内 章『e-Learning 教材への下線引きによる読解支援に関する研究』電子情報通信学会技術研究報告. ET, 教育工学 105(47), 55-60, 2005-05-07

[10] 鈴木優, 布目光生, 石谷康人『インタラクティブなペン操作を可能とする検索意図に基づく連鎖情報検索』情報処理学会シンポジウム論文集 20060302 巻: 2 0 0 6 号: 4 頁:101-108