小学生を対象としたメッセージ推奨のためのキーワード提示システム

須田 幸次†† 永田 亮† 掛川 淳一† 森広浩一郎† 正司 和彦†
† 兵庫教育大学 †† 神戸市立下緑小学校
E-mail: †{m05321b,nagata,kakegawa,mori,showji}@hyogo-u.ac.jp

1. はじめに


このような背景を受け、須田ら[3]は、小学生を対象とした情報発信に関する学習を支援するシステムを提案している。このシステムでは、児童が自分たちで考えた本について「おすすめメッセージ」と呼ばれる本の推薦文をログの記事として情報発信する。小学校 5 年生を対象とした実験の結果、児童は積極的で学習に取り組み、頻繁に情報発信を行うことが確認されている。

一方で、須田ら[4]は、児童が発信する情報の質や大きな問題が残っていることも報告している。児童は内容がどうとんどない情報を発信する傾向が非常に多い、典型例として“おもしろい本です”や“楽しい本です”などが多く挙げられる。理想的には、本の概要や面白さにどんな情報が含まれているかを推奨されるべきである。

この問題に対する直接的なアプローチとして、教師が情報発信の内容を指導することが考えられる。児童一人一人に対して情報発信の手本を提示する指導である。しかしながら、このような場合、個別に指導を行うため、教師の負担が大きくなる。特にこの場合、児童が提示された例を単に書き写すという活動に終わってしまうと、より高い学習効果をあげるためには、児童が、キーワードとなる適切な語を選定し、受け手に対して価値のある情報となるよう情報発信の内容を推奨することが重要である。

そこで、本論文では、キーワードを提示することで発信内容の推奨を児童に促す学習支援システムを提案する。例えば、「桃太郎」を読んで「おもしろい」と発信した児童に対して、キーワード「桃、きだんご、鬼」などを提示するシステムである。提案システムにより、児童は(1) 適切なキーワードの選択、(2) キーワードからの本の内容の思い出し、(3)「おすすめメッセージ」の推奨の三種類の活動を行うことになり、学習効果が期待できる。提案システムは、本のタイトルと児童が作成した「おすすめメッセージ」に合わせて、キーワードの提示を行う。また、提案システムは、提示の候補となるキーワードの選択を助ける。したがって、キーワード提示に係る教師負担は無く、多数の児童が頻繁に情報発信する環境でも学習支援が可能となる。

以下、2.で、提案システムの基本アイデアを述べる。3.で、提案システムの詳細を説明する。4.で、提案システムを利用した評価実験について述べ、その結果を考察する。

2. 基本アイデア

提案システムは、ブログを利用した学習支援システム[6]上に実装される。学習支援システム[6]は、児童が選択した本のタイトルや著者などの書籍情報を自動的にブログの記事として記入する(本のタイトルが、記事のタイトルとなる)[注1]。読書後、児童は、その本の「おすすめメッセージ」をブログの記事に記入する。例えば、「桃太郎」を読んだら、「おもしろい話です」と書いたとする。このとき、提案システムは、本のタイトル、及び「おすすめメッセージ」中の単語を基に、適当なキーワードを児童に提示し、「おすすめメッセージ」の推奨を促す。児童は、提示されたキーワードから適切なキーワードを選択し「おすすめメッセージ」を推奨する。最終的に、推奨した「おすすめメッセージ」の発信を行う。児童は、この活動を繰り返すことにより情報を発信する学習を行う。

提示するキーワードの候補は、過去に発信された全ての児童の記事から自動的に収集する。収集のための知識源は、記事のタイトル(本のタイトルでもある)と「おすすめメッセージ」である。さらに、学習支援システム[6]では、本の帯情報も利用可能であるため知識源とする[注2]。帯情報とは、本の巻に書かれた宣伝文のことである。ただし、帯情報が取得できない本に関しては利用しない。同様の一の本を、複数の児童が読む(注1): 本研究では、神戸市立中央図書館(http://www.city.kobe.lg.jp/library/main/index.html)のデータにより、インターネット経由で書籍情報を得ている。このサービスは、神戸市内の公立学校・中学校であっても受けることができる。このサービスが受入れられない場合でも、一般的の受信機から図書情報を取り扱うことができる。

(注2): 帯情報については、神戸市立中央図書館の協力を得ている。なお、帯情報も書籍情報として扱われる可能性がある。
可能性があるので、「おすすめメッセージ」も一冊の本に対して複数存在することがある。また、「おすすめメッセージ」が未記入の記事もある。したがって、1つの知識源は、本のタイトル、0以上の「おすすめメッセージ」、0または1つの都情報から成る。以下では、この知識源から自動収集したキーワード候補を登録した辞書、キーワード辞書と呼ぶことにする。
提案システムは、各キーワード候補に対してスコアを計算し、そのスコアに基づいてキーワードの提示を行う。スコアの直感的な解釈は、

スコア = 重要度 x 関連度 (1)

で、与えられる。すなわち、重要度も関連度も高いキーワード候補が高いスコアを得ることになる。重要度とは、キーワードとしての適切さを表す。例えば、「本」という語は、大部分の文で意味を有しているわけではない。一方、「しあわせだね」という語は、特定の文の中で出現し、本を特徴付ける重要度の高いキーワードといえる。また、関連度とは、「おすすめメッセージ」とキーワード候補の関連の度合いである。例えば、「西遊記」において「おすすめメッセージ」を記述する際に、「鬼、桃」などを提示しても関連度が低いため良いキーワードとはいえないと、「西遊記」に対しては、関連度が高い「石、悟空」などを提示すべきである。

重要度と関連度の定量化には、様々な手法が考えられるが、提案システムでは、次の二点を考慮して定量化を行う。第一に、単純な手法であることを重視する。提案システムは、主に小学校で使用されるものを考慮すると、それほど高性能ではない計算機に安価に実装できる必要がある。また、システムに対する有効なフィードバックを教師から得るためには、単純な手法であることが好ましい。第二に、必ずしも高い精度でキーワードを提示する必要はないと考える。なぜなら、たとえ不適切なキーワードが含まれて提示されたとしても、その中から適切なキーワードを選択するという児童の学習活動を繰り返すからである。

以上が、提案システムの基本アイデアである。次章で、提案システムの詳細について述べる。

3. 提案システム

3.1 キーワードのスコア計算

スコアの計算方法を定式化するための記号を導入する。

N 個の知識源、B_1、B_2、…、B_N に与えられており、これらの知識源から全部で M 個のキーワード候補 w_1、w_2、…、w_M が得られたとする（本稿では、特に断らない限り、この知識源を単に本と呼ぶ）。また、キーワード候補 w_i を含む本の数を n_i で表す。

いま、ある児童が「おすすめメッセージ」を書き、その本文またはタイトルに単語 w_j が出現したとする。このとき、キーワード候補 w_i と w_j との関連度を相互情報量

\[ r(w_i, w_j) = \log \frac{P(w_i, w_j)}{P(w_i)P(w_j)} \] (2)

で定義する。ただし、P(w_i) と P(w_i, w_j) は、それぞれ w_i の生起率と w_i と w_j の共起率である。生起確率と共起確率は、知識源である本の集合から推定する。次に、キーワード候補 w_i の重要度を IDF

\[ a(w_i) = \log \frac{N}{n_i} \] (3)

で定義する(注3)。更に、式 (1)、式 (2)、式 (3) を用いて、「おすすめメッセージ」、またはそのタイトルに単語 w_j が出現したときのキーワード候補 w_i のスコアを、

\[ s(w_i, w_j) = a(w_i) \times r(w_i, w_j) \] (4)

で定義する。式 (4) の 1 つの解釈として、「おすすめメッセージ」に出現した単語 w_j が、スコア（確信度）s(w_i, w_j) で、キーワード候補 w_i を推定する考えることができる。

一般に、「おすすめメッセージ」とそのタイトル中には、複数の単語が出现する。したがって、これらの単語から計算される全ての s(w_i, w_j) を考慮して、キーワードを提示する必要がある。そこで、提案システムでは、s(w_i, w_j) の重みつき和をとり最終的なスコアとする。特に、キーワード候補の推奨元となる単語 w_i の重要度とする。すなわち、推奨元となる単語が重要であるほど、最終的なスコアに対する影響が大きいとする。以上より、最終的なスコアを、

\[ S(w_i) = \sum_{w_j \in W} a(w_j) s(w_i, w_j) \] (5)

で計算する。ただし、W は対象とする「おすすめメッセージ」及びそのタイトル中に出現したキーワード候補の集合を表す。また、a(w_j) は、w_j の本中での出現頻度を表す。
提案システムでは、「おすすめメッセージ」のタイトル（すなわち、児童が読んだ本のタイトル）中の単語を優先的に提示するため、タイトル中の単語には高いスコアを与える。具体的には、タイトル中のキーワード候補 w_i に対しては、

\[ S(w_i) = \sum_{w_j \in W} a(w_j) s(w_i, w_j) + \alpha \] (6)

で、スコアの計算を行う。ただし、\( \alpha \) は、タイトル中の単語に対する優先度を決定するパラメータである。また、タイトル中の単語でキーワード辞書に登録されていない単語（ここでは \( \hat{t} \) とする）については、

\[ S(t) = \alpha \] (7)

(注 3)：関連度と重要度の解釈を容易にするため、0 ≤ r(w_i, w_j) ≤ 1、0 ≤ a(w_i) ≤ 1 となるよう正規化を行う。なお、正規化してもキーワード候補間の重要度・関連度の大小関係は変わらない
で、スコアを計算する。提案システムでは、4.で述べる予備実験の結果から \( \alpha = 0.3 \) とする（以下の述べるシステムのパラメータについても、この予備実験の結果を基に決定した）。

以上が、提案システムで用いるキーワード候補のスコア計算方法である。重要度と関連度の定義に示す。上述の方法以外にも様々な方法が考えられる。例えば、重視度として IDFを用いたが、IDFだけでも RDF [6] など様々なバリエーションがある。関連度についても、語の類似度 [3] などを利用できる。これらの手法は、提案システムに比べ、複雑な計算を要したり従来の知識を利用したりするため、より高い精度でキーワードの指示が行えると予想される。しかしながら、提案システムでは、2.で述べた設計思想に基づき、比較的単純な手法で重要度と関連度を定量化する。

3.2 キーワード辞書の自動構築

キーワード辞書の自動構築は、以下のステップから成る。

(ii) キーワード候補の抽出、(ii) スコア算出、(iii) キーワード候補の保存。

(i) では、本の集合からキーワード候補を抽出する。まず、形態素解析を用いて、知識源である本を単語に分割する。本論文では、形態素解析ツール「Kawari」 [4] を用いる。次に、分極した語から、名詞が、名詞一覧、名詞一覧を基に、形態素解析を用いて、キーワード候補として抽出する。スコア算出に品詞情報も「スコア」の出力を利用する。ただし、連続する名詞は、連続し1つのキーワード候補として扱う。また、二つの名詞の間にカタカナ、カタカナ、記号が含まれる場合は抽出の対象とし、更に、低頻度なキーワード候補を、得られた統計数の信頼性が低い場合、頻度3以下の語はキーワード候補から除外する。

(ii) では、式 (4) を用いてスコアの計算を行う、スコアの算出、キーワード候補の全ての組み合わせについて。更に、低頻度なキーワード候補を、得られた統計数の信頼性が低い場合、頻度3以下の語はキーワード候補から除外する。

(iii) では、キーワード候補を保存してキーワード辞書を構築する。また、推奨元となる単語、その語の重要度、スコアも保存する。

3.3 キーワードの提案による学習支援

キーワードの提案による学習支援は、以下のステップから成る。（i）「おすすめメッセージ」からの単語の抽出、（ii）スコアの算出、（iii）キーワード学習支援メッセージの提示。

(i) では、キーワード候補から、おすすめメッセージを基に、そのタイトルから単語を抽出する。抽出した単語は、3.2のステップ(i)と同様である。ただし、頻度3以下の語は抽出する。

(ii) で、抽出した単語に基づき、キーワード候補の最終的なスコアを計算する。最終的なスコアの計算に必要となる情報は、キーワード辞書に登録されていないキーワード候補、推奨元の単語については重要度及びスコアを0と見なす。

(iii) では、計算したスコアに基づきキーワードを提示し、学習を支援する。スコアの高い順に、キーワード候補を3つ提示する。ただし、「おすすめメッセージ」には含まれなる語は提示の対象とはしない。また、パラメータ \( \alpha \) よりスコアが低い語も対象としない。キーワードと共に「おすすめメッセージ」の推奨を促すシステムメッセージを提示し学習を支援する。例えば、「桃太郎」を読んだで「楽しい本です。」と書いた児童に、「キーワード「桃太郎」と共に、システムメッセージ「これらのキーワードを使って、おすすめメッセージを書き直してみよう。」と提示して推奨を促す。その結果を受けて、児童は推奨を行う、複数回の推奨が行えるように、推奨した「おすすめメッセージ」に対しても、再度(1)〜(3)の手順を行いキーワードの提示を行う。どのページで推奨を終了するかは、児童の判断に任せて行うこととする。

4. 評価実験

本章では、提案システムを評価するために実施した実験について述べる。評価は、キーワードの提示精度と「おすすめメッセージ」の質の変化との2点について行った。

4.1 実験条件

神戸市立X小学校5年生（3学級98名）を対象として、提案システムの評価を行った。各学級3級（学級65〜65）の89〜85の実践を行った。各学級とも、提案システムのベースとなる学習支援システム [5] を事前に8級使用した。

4.2 実験手順

まず、キーワード辞書を構築した。学習環境システム [6] の事前8級の使用から得られたデータを知識源とした。

次に、予備実験を行った。予備実験は、各学級1級の8級実施した。予備実験の主な目的は、3で説明したパラメータの調整である。また、児童に提案システムの使用方法を説明した。更に、提示されたキーワードを言葉に「おすすめメッセージ」を推奨する実演を行った。

予備実験後、本実験を実施した。評価用に、ブログの記事ID, 書き換え前の「おすすめメッセージ」、本のタイトルを提案したキーワード、書き換え後の「おすすめメッセージ」、書き換え時間を一覧の情報としてログとして記録した。

最後に、キーワード提示の精度と「おすすめメッセージ」の質の変化を評価した。キーワード提示の精度は、第一著者と第二著者が一人で評価した。提示されたキーワードに対して、「適切」、「不適切」、「どちらともいえない」のいずれかで評価した。ただし、「書き換え時間を除外」は同一であるログ中のみのキーワードに対しては一度のみ評価した。評価尺度は、「適切」、「不適切」、「どちらともいえない」のいずれかで評価した。評価尺度は、「適切」、「不適切」、「どちらともいえない」のいずれかで評価した。評価尺度は、「適切」、「不適切」、「どちらともいえない」のいずれかで評価した。評価尺度は、「適切」、「不適切」、「どちらともいえない」のいずれかで評価した。
るか判定してもらった。ただし、複数回書き換えが行われている場合は、書き換え前の最後の書き換え結果をベースにして提示した。

4.3 評価結果と考察
表1に、キーワード提示精度の評価結果を示す。表1中の"Title","General","Total"は、それぞれタイトルから得られたキーワード、タイトル以外から得られたキーワード、全てのキーワードを対象にしたときの精度を意味する。

表1から、提案システムは、高い精度でキーワードを提示できたことがわかる。特に、タイトルから得られたキーワードの精度が高い。タイトル以外から得られたキーワードについても高い精度を達成しており、全体の精度は65%を越えている。このことは、システムが提示する3つのキーワードのうち、平均的に2つは適切なものであるということを意味する。言い換えると、提案システムから、適切なキーワード2つを選び出し、「おすすめメッセージ」を推奨することになる。評価者間の一致度を表すk値は0.440となり、中程度の一致度を示した。

次に、「おすすめメッセージ」の質の変化に対する評価について述べる。実験中に、131の情報発信が行われ、そのうち29の記事（22%）について「おすすめメッセージ」の書き換えが行われた。一見のこの数値は低く見えるが、授業時間が45分であることを考慮すると妥当な数値であると考えられる。45分の間に、児童は本の選択、読書、「おすすめメッセージ」の記述、「おすすめメッセージ」の発信、提示されたキーワードを利用した「おすすめメッセージ」の推奨、再発信の全てを行わなければならない。さらに、児童三人あたり一台のパソコンを利用していたことも考慮しなければならない。

一方で、書き換えが行われた「おすすめメッセージ」について、目覚ましい改善が見られた。書き換え前の「おすすめメッセージ」は平均19.9文字であったのに、書き換え後は平均35.7文字まで増加した（有意水準99%で有意、paired t-test）。このことは、提示されたキーワードの大部分が2文字〜3文字から成ることを考えると、児童は提示されたキーワードだけでなく、自分自身の言葉を加えて「おすすめメッセージ」を推奨していることがわかる。質的評価でも、それは目的の評価者は、26の「おすすめメッセージ」のうち21が書き換え後のほうが良いと評価した（残り3つについては、同程度良いと評価した）。同様に、二番目の評価者は、23のうち22が書き換え後のほうが良いと評価した（残り6つは同程度良いと評価した）。したがって、のべの改善率は、88%（43/49）となる。実際、児童は、提示されたキーワードから適切なものを選択して、「おすすめメッセージ」をより良いものに書き換えている。例えば、「西遊記」を読んで“いやすごい、きっかけで。"と書いた児童は、「この市場に書かれている"と書き換えている。この例は、語の分割誤り（"悟空"→"悟空"）を、児童が適切に修正することを示している。

5. おわりに
本論文では、キーワードの提示により、情報発信の学習を支援するシステムを提案した。実験の結果、次3の点が確認できた。第一に、比較的単純な手法でも十分な精度でキーワードを提示できる。第二に、キーワードの適切な提示には、情報発信の質を高める学習効果がある。第三に、学習効果を上げるためには、必ずしも高い精度でキーワードを提示する必要はない。

今後の課題として、より頻繁に「おすすめメッセージ」の書き換えが行われるように、キーワードの提示方法を改善することが挙げられる。また、提案システムを利用した指導方法の改善も挙げられる。

謝辞
本研究にあたって、書誌データを提供いただいた神戸市立中央図書館に感謝いたします。また、評価実験に協力をいただいた大田市立大田小学校、有馬小学校教員に感謝いたします。

参考文献