

Web ニュースサイトのランキング情報を用いた 小学生のための Web ニュースの推薦

Web News Recommendation using Ranking Information on Web News Sites for Elementary School Students

田中 翔也*¹
Shoya Tanaka

安藤 一秋*¹
Kazuaki Ando

*¹ 香川大学工学部
Faculty of Engineering, Kagawa University

Recently, NIE (Newspaper In Education) has attracted attention at elementary schools. The use of newspaper for education has been specified in the government course guidelines from the fiscal year 2011. However, there are kanji characters or terms which students cannot read and understand in newspapers. The aim of our research is to construct a Web news selecting support system for elementary school students in order to enhance the effectiveness of NIE. This paper proposes a recommendation method of newspaper articles that might interest elementary school students by ranking Web news based on news ranking on Web news sites, query logs containing trend words, and learning knowledge.

1. はじめに

近年、小学校では、新聞を教材として活用する教育 (NIE: Newspaper in Education) が実施されている。特に、平成 23 年度から実施されている新教育指導要領においては、新聞を授業で利用することが明示され、NIE の重要性は益々高まっていると考えられる。しかし、一般の新聞や Web ニュースは小学生を対象として書かれておらず、小学生が学習していない漢字や表現などが使われる。したがって、小学生にとって記事の内容を理解することは難しい。

NIE では、一般的に紙媒体の新聞を利用するが、近年では、新聞を購読しない家庭も増えている。新聞を購読していない家庭においては、Web ニュースを利用して実施することになる。しかし、小学生は語彙力および検索力が乏しいため、十分な検索クエリを設定することは容易ではない。そのため、膨大な記事の中から自身の興味や授業テーマに合った記事を探し出すことは難しい [坪井 13]。

本研究では、小学生が興味をもつ、あるいは興味を持ってもらいたい記事を推薦する手法の実現を目的とする。本稿では、複数の Web ニュースサイトにあるランキングに注目する。ニュースランキングは、ニュースの注目度を表す情報源であるため、小学生が興味をもつ、あるいは興味を持ってもらいたい記事が埋もれている可能性が高い。しかし、ニュースランキングは、小学生を対象としたランキングではないため、NIE における記事推薦のランキングとして、そのまま利用することは難しい。そこで、子供の興味や埋もれていると考えられる、子供向けポータルサイトのキーワードランキングに注目し、そこから得られる言葉 (トレンド言葉) の活用を検討する。また、教科書に含まれる重要語や単元情報、学習支援サイトに含まれる重要語などの学習知識の利用も検討する。

本稿では、当研究室で構築を進めている NIE 支援システムと関連研究について述べた後、推薦手法の基本方針と、実装に向けた 2 つの事前調査について述べる。

2. 当研究室の NIE 支援システムの概要

当研究室では、小学校での NIE を対象とした NIE 支援システムの構築を進めている。本システムでは、新聞記事の検索・選出支援、読解支援、調べ学習支援の 3 つの支援の実現を目指している。また、本システムの対象ユーザは、小学校高学年の児童と教師である。

新聞記事の検索・選出支援として、坪井ら [坪井 13] は、小学生と教師を対象とした記事選出支援を提案している。この研究では、小学生が単独で新聞記事を読む際、不適切となる記事のフィルタリング手法を提案している。また、教師の負担となっている記事の選出の自動化及び記事推薦を教科書と記事の類似性判断と適性フィードバックを用いて実現している。

読解支援として、藤沢ら [藤沢 12] は、Web 上の一般向け新聞記事を子供向けに言い換えるための知識を構築する手法を提案している。具体的には、Web 上の一般向け新聞記事とそれに対応する子供向け新聞記事のペアを情報源とし、係り受け関係を利用して言い換え表現対を抽出する。また、小林ら [小林 12] は、Web 上の新聞記事から難易語や重要語を自動抽出する手法を実現し、それらに対する説明の付与、記事に関連するデータや写真の提示、複数の新聞記事の比較を行うことができる読解支援システムを提案している。

調べ学習支援として、久保ら [久保 13] は、ユーザの理解度を考慮しながら調べ学習支援を行うシステムを提案している。この研究では、調べ学習課題の提示、参考資料の提示、フィードバックにより、小学生の新聞記事内容に関する調べ学習を支援することを目的としている。また、記事に含まれる重要語を基に調べ学習課題を自動生成する手法を提案している。

本研究では、新聞記事の検索・選出支援として、小学生が興味をもつ、あるいは興味を持ってもらいたい記事を推薦する手法の実現を目指す。

3. 関連研究

近年、インターネットの普及と共に情報が溢れ返っている。それに伴い情報選択の必要性が生じ、フィルタリングの需要が高まっている。このような問題は一般人だけではなく、語彙が乏し

く、クエリを決定しにくい子供に対しても大きな問題である。この問題を解決するため、子供に対する Web 検索支援や情報整理の研究が行われている。

岩田ら[岩田 13]は、サブピックに注目し、子供向けの Web 検索結果のランキング手法を検討している。サブピックとは、入力されたクエリに対してユーザの意図を具体化する文字列のことである。例えば、“北海道”というクエリの場合、“旅行”、“スキー”などがサブピックの候補に挙げられる。しかし、北海道に対しては、このようなサブピックが一般的に重要であると考えられるが、これらは子供にとって興味を持ちやすいサブピックではない。一方、“食べ物”、“歴史”などのサブピックは、子供が授業で北海道の文化や習慣について学んでいる際に興味を持ちやすい内容であり、子供にとって重要であると考えられる。岩田らは子供用検索サイトのクエリに対する関連検索語や検索結果上位のスニペットに出現するキーワードに注目し、そこから抽出したサブピックを利用して子供にとって有効なランキングを行う手法を検討している。検索目的に応じてどのようなサブピックが有効であるかを調査した結果、目的により重要なサブピックが変化すると述べている。また、学習に関するクエリは、子供にとって一見親しみにくくても、役に立ちそうなサブピックが有効であると述べている。

本研究では、学習を目的とした記事検索を行うため、サブピックの設定も容易であると考えられる。サブピックの利用については今後検討する。

西ら[西 95]は、ユーザ情報と記事情報の 2 つの情報のマッチングによりランキング値を算出し、その値で新聞記事をランキングする手法を提案している。記事情報は、意識的な情報としてユーザが入力したクエリログ、ユーザ情報は、無意識的な情報としてユーザが入力した年齢、性別、職業、出身などの個人情報である。意識的な情報とは、記事中のキーワードに対する関心度を定義し、見出しや本文における出現頻度と合わせて定義する。意識的な情報と無意識的な情報を合わせ、ランキング値として算出している。

キーワードの関心度計算は、本研究における、トレンド言葉や学習知識などに対する重み付けの参考にできる。

4. 記事推薦システムの概要

現在検討している記事推薦システムについて説明する。ニュースランキングは、ニュースの注目度を表す情報源であるため、小学生が興味をもつ、あるいは興味を持ってもらいたい記事が埋もれている可能性が高い。しかし、ニュースランキングは、小学生を対象としたランキングではないため、小学校の NIE で利用する記事ランキングとして、そのまま利用することは難しい。そこで、小学校の NIE に適したランキングにリランキングする必要がある。

本稿では、リランキングに利用する情報として、まず、以下の 3 つに注目する。

- 1) 複数のニュースランキング
- 2) 小学生向けポータルサイトなどのキーワードランキングから抽出したトレンド言葉
- 3) 教科書や学習支援サイトから抽出した学習知識

複数のニュースランキングを利用することで、より注目されている記事を把握できる。また、子供の興味や人気のある言葉を収集するため、子供向けポータルサイトのキーワードランキングに注目する。キーワードランキングに含まれる言葉は、子供達の現在の興味を表していると考えられる。そこで、子供向けサイトのキーワードランキングから特徴的な言葉を抽出し、リランキングに活用する。以降、キーワードランキングから抽出した特徴的

な言葉を「トレンド言葉」とよぶ。さらに、小学校での NIE に適したランキングを実現するため、教科書に含まれる重要語や単元情報、学習サイトに含まれる重要語などの学習知識をリランキングに利用する。

ニュースランキングを活用した記事推薦システムのイメージを図 1 に示す。

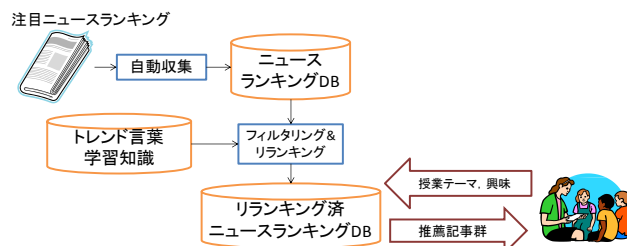


図 1. 記事推薦システムのイメージ

本システムでは、まず、複数の新聞社サイトからニュースランキング情報を収集し、データベースに格納する。そして、トレンド言葉や学習知識を基に、収集したランキング情報をリランキングする。この情報を基に、ユーザに対して注目ランキング記事群を推薦する。

5. リランキング手法の基本方針

現在検討しているリランキング手法の基本方針について説明する。リランキングは、ランキング情報、トレンド言葉、学習知識の 3 つの情報を用いる。

ランキング情報による重み付けでは、複数の新聞社サイトから得たランキング情報を用いる。複数の新聞社サイトのランキングには、同じ対象を扱った記事が存在する可能性が高い。しかし、これらの記事に同じ見出しが使われる保障はない。新聞社によって異なる見出しが付けられたり、本文の内容が若干異なる可能性もある。このような記事を異なる記事と見なすと、人気がある記事でも上位に上がらない場合がある。逆に、同じ記事がランキングに何件も含まれてしまう可能性もある。そこで、記事見出しやリード文の類似度を基に、同一対象を扱った記事を判定する。続報記事に対する処理については、今後検討する。

トレンド言葉による重み付けでは、Yahoo!きっずの「みんながよく調べる言葉ベスト 20」とキッズ goo のにんぎのこことばページの「今週のランキング」、学研キッズネットの「人気のキーワード」と「注目キーワード」を利用する。なお、キーワードランキングには、新聞学習に有用ではないキーワードを含んでいる可能性があるため、それらをフィルタリングする必要がある。また、利用するキーワードランキングも今後検討が必要である。

学習知識による重み付けでは、教科書の重要語、単元情報や学習サイト(キッズ goo のしらべものノートなど)の重要語などを調査し、どのようなキーワードが利用できるかを検討する。

以上の 3 つの重みを利用した計算式についても、今後検討する。

6. トrend言葉に対する調査

トレンド言葉がリランキングの重みとして利用できる可能性を確認するために、トレンド言葉が記事に出現する含有率を調査する。調査に利用するトレンド言葉は、Yahoo!きっずの「みんながよく調べる言葉ベスト 20」とキッズ goo のにんぎのこことばページの「今週のランキング」、学研キッズネットの「人気のキーワード」と「注目キーワード」から収集した 46 語(各 20 語, 10 語, 16

語)である。本来であれば、これらのランキングから特徴的な語を抽出して利用すべきであるが、本調査ではそのまま利用する。調査対象の記事は、2013年11月1日から2014年1月22日までの記事326件を利用する。また、調査方法は、トレンド言葉および新聞記事を形態素解析し、形態素単位の一貫で判定する。

調査の結果を表1に示す。表1より、キッズgooサイトから収集したトレンド言葉の含有率が40.2% (131/326)と最も多く、半分近くの記事に含まれていた。トレンド言葉は、子供が興味を持って使った単語と考えられるため、トレンド言葉が含まれている新聞記事は興味のある記事である可能性がある。そのような記事に対して、重み付けを行うことで、子供が興味を持ちやすい記事の推薦が可能であると考えられる。しかし、トレンド言葉を含む記事が毎日出現するとは考えにくいので、重みづけを含め、さらなる調査が必要であると考えられる。

表1. トレンド言葉の含有率の調査結果

使用サイト	含有率(%)
Yahoo!きっず	17.2
キッズgoo	40.2
学研キッズネット	18.7

それぞれのサイトから得られた出現記事数が上位3件のトレンド言葉を表2に示す。表2からもわかる通り、1文字からなる単語が上位に出現している。この場合、記事中の他の単語とミスマッチングしたことも考えられる。トレンド言葉は、フレーズの場合もあるが、本調査では、形態素解析により単語に分割して調査したため、フレーズでの調査も必要である。また、小学校でのNIEには不要である言葉も含まれる可能性があるため、今後は、フィルタリングを検討する必要がある。

表2. 各サイトのトレンド言葉上位3件

Yahoo!きっず		キッズgoo		学研キッズネット	
単語	記事数(記事)	単語	記事数(記事)	単語	記事数(記事)
人	20	国	39	世界	23
冬	7	人	20	時代	12
米	7	地方	16	自由	8

7. 教科書の見出しに対する調査

教科書の章・節・項の見出しがリランキングの重みとして利用できる可能性を確認するために、見出しを構成する名詞(見出し語)が記事に出現する含有率を調査する。

平成22年の小学6年生の社会・下(東京書籍)の教科書に存在する、章・節・項の見出しを構成する名詞(不要語を除く)37語を見出し語として利用する。調査対象記事は、6の調査と同じものを利用する。調査方法は、新聞記事を形態素解析し、見出し語と記事の形態素単位の一貫で判定する。

調査の結果、含有率は25.8% (84/326)となり、約1/4の記事に含まれることがわかった。また、見出し語の記事頻度を降順に並べた結果を表3に示す。本調査では、複合名詞化せずに頻度を数えたため、抽象的な語が上位に存在している。

複数の見出し語が多く含まれている新聞記事は、教材として教師が学んでほしい内容を多く含んでいると考えられ、NIEに用いる新聞記事として有用であることが予想される。しかし、本調査は、少ない単語で調査であるため、教科書全体から抽出した学習知識を利用した調査を行う必要がある。

表3. 学習知識の調査結果

キーワード	記事(数)
生活	16
国会	13
センター	10
まち	8
公共	8
実現	8
政治	7
選挙	6
願い	2
生涯	2
学習	2
税金	1
大切	1
合計	84

8. おわりに

本稿では、複数の新聞社サイトのランキング情報を基に、トレンド言葉と学習知識を利用してリランキングする手法について検討した。Yahoo!きっずとキッズgoo、学研キッズネットのキーワードランキングからトレンド言葉を収集し、その言葉が新聞記事に含まれる含有率の調査した結果、トレンド言葉が17%から40%程度の新聞記事に出現していた。また、教科書の章・節・項の見出しを収集し、見出しに含まれる名詞が新聞記事に含まれる含有率を調査した結果、25.8%であった。本調査では、利用する見出し語の数が少ないため、教科書全体から抽出した学習知識を利用した調査を行う必要がある。また、キッズgooのしらべものノートなどの学習支援サイトに含まれる重要語の利用についても、今後検討する。今後の課題として、事前調査を継続する。また、既存手法[岩田13]も参考に、サブトピックの利用についても検討する。そして、ランキング情報、トレンド言葉や学習知識に対する重み付けを行う手法を考案する。最終的に、記事推薦システムとして実装し、小学校で実験・評価を行う予定である。

謝辞

本研究の一部は、JSPS 科研費 25350335 の助成を受けて実施した。

参考文献

- [坪井 13] 坪井賢泰, 安藤一秋: 小学生向け NIE を対称とした Web 新聞記事の推薦, 情報処理学研究報告, コンピュータと教育研究会報告, 2013-CE-119, 2013.
- [藤沢 12] 藤沢祐輔, 相原慎太郎, 安藤一秋: Web 一般新聞記事を子供向けに言い換える知識の抽出, 言語処理学会第18回年次大会論文集, p.751-754, 2012.
- [小林 12] 小林健, 藤沢祐輔, 坪井賢泰, 安藤一秋: 小学生を対象とした Web 新聞記事読解支援のための重要語抽出, 言語処理学会第18回年次大会論文集, pp.843-846, 2012.
- [久保 13] 久保恵津子, 安藤一秋: 新聞読解支援における調べ学習課題自動生成の検討, 2013年度人工知能学会全国大会(第27回), pp.1-3, 2013.
- [岩田 13] 岩田麻佑, 原隆浩, 西尾章治郎: Web 検索クエリに関する子供向けサブトピックランキング手法の検討, 日本データベース学会, DEIM Forum 2013 A1-5, 2013.
- [西 95] 西孝史, 中嶋卓雄, 北川結香子, 河北隆生, 中村良三: ユーザの好みを考慮した新聞記事のランキング, 情報処理学会第55回全国大会, pp.3-228-3-229, 1995.