

SNS での注目度と教育的知識に基づいた
小学生のための Web ニュースランキング手法
A Web News Ranking Method based on Degree of Attention of SNS Users and
Educational Knowledge for Elementary School Children

田中 翔也†
Shoya Tanaka

安藤 一秋‡
Kazuaki Ando

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[1]. NIE では、一般的に紙媒体の新聞を利用しているが、新聞を購読しない家庭の増加に伴い、Web ニュースを利用する機会も多くなった。しかし、小学生は語彙力と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。紙媒体の新聞と Web ニュースを問わず、膨大なニュース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[2]. また、教師にとっても、NIE に適した記事を探し出すことやその時間確保も難しい状況である。そこで、NIE に適した新聞記事を推薦できれば、この問題を改善できると考える。

ニュースサイトには、ニュース記事のアクセス回数に基づくアクセスランキングが存在するが、このランキングには、読者が記事を読んで興味を持ったか否かの情報が反映されていない。そこで本研究では、SNS に注目する。YOMIURI ONLINE や産経ニュースなどのニュースサイトには、各記事ページに SNS と連携する機能があり、読者が記事を読んだ後に興味をもてば、SNS に対して簡単に投稿できる。さらに、SNS の代表例である Twitter は、ツイートの 85% がニュース記事に関連するという報告[3]もある。したがって、SNS での注目度が高いニュース記事は、小学生が興味を持つ、あるいは興味を持ってもらいたい記事である可能性がある。しかし、SNS での注目度のみを利用して生成されたランキングは、必ずしも NIE に適したランキングとはいえない。

そこで本研究では、SNS での注目度だけでなく、子供向けポータルサイトにおいて人気のある検索語、教育的知識 (教科書の重要語・単元情報、教師が指定するキーワード) などを総合的に利用して、小学生が興味を持ちやすい、あるいは興味を持ってもらいたい記事を推薦するシステムの構築を目的とする。

本稿では、まず、ニュースサイトに掲載されているアクセスランキングとソーシャルランキングの傾向を分析する。そして、教育的知識の予備調査として、社会の教科書の見出しに注目し、見出しから抽出した単語がニュース記事に含まれる割合を調査することで、教育的知識として利用できる可能性について考察する。

2. ニュース記事のアクセスランキングと SNS の注目度に基づくランキング

ニュースサイトには、ニュース記事のアクセス回数に基づくアクセスランキングが存在する。アクセスランキングは、単純にアクセス数を基にしているため、読者が記事を読んだ後、その記事に対して興味を持ったか否かの情報が反映されていない。

一方、SNS での注目度に基づくランキングの場合、興味のない記事に対しては、SNS に投稿される可能性は低いと考えられる。したがって、SNS での注目度を利用することで、記事を読んで多くのユーザが興味もった記事のランキングが生成できると考えられる。

この仮定の妥当性を確認するため、次章では、ニュースサイトに掲載されているアクセスランキングとソーシャルランキングの特徴を分析する。

3. アクセスランキングとソーシャルランキングの傾向分析

アクセスランキングとソーシャルランキングの特徴を分析するためには、同じ記事集合を扱うランキングを比較対象に利用することが望ましいと考える。

そこで、複数のランキングが掲載されている産経ニュースサイトのアクセスランキングとソーシャルランキングを分析対象として利用する。なお、産経ニュースサイトのアクセスランキングは、記事に対するアクセス数を基に生成され、ソーシャルランキングは Twitter と Facebook に投稿された数の合計値が基に生成されたものと考えられる。

3.1. データ

分析には、2015年6月24日から6月26日までの3日間に収集したランキングデータを利用する。各ランキングは、それぞれ30位までの順位情報が利用できるため、1日60件、全180件のランキングデータを分析に利用する。

分析方法として、アクセスランキングとソーシャルランキングに含まれる記事の重複率を確認する。また、二つのランキングの両方あるいは片方にしか存在しない記事の特徴を確認する。

3.2. 結果と考察

アクセスランキングとソーシャルランキングの記事の重複率を確認した結果、1日30件のランキングの内、平均して51.0% (15.3記事) がランキングに重複して存在することを確認した。この結果から、各ランキングの約半分が異なる記事で構成されていることがわかる。また、ランキングの順位が重複している記事は、平均して2.2% (0.6件) であった。このことから、同じ記事でもほとんどの記事で順位が異なることがわかる。

† 香川大学 大学院工学研究科
Kagawa University Graduate School of Engineering

‡ 香川大学 工学部
Kagawa University Faculty of Engineering

次に、アクセスランキングとソーシャルランキングの両方に存在する記事と、アクセスランキングにしか存在しない記事の傾向を調査した。調査の結果としては、あまり記事の傾向に大きな差は確認されなかった。特に、ソーシャルランキング 30 件を 10 件ずつの区間に分けて注目すると、表 1 に示す結果となった。

表 1 ランキングの順位と重複率の関係

区間	平均重複件数 (件)	平均重複率 (%)
1~10 位	6.00	60.0
11~20 位	3.00	30.0
21~30 位	0.67	6.7

表 1 より、ランキングの順位が下位になるほど、重複率が減少していることがわかる。また、この関係は比例関係にあることも確認できる。上位 10 件でも平均 4 件は記事が重複していないことを踏まえると、各ランキングの利用者層が若干異なることも推測される。

分析結果をまとめると、アクセスランキングとソーシャルランキングでは、約半分の記事が異なる。また、片方のランキングにしか存在しない記事に大きな特徴は見られない。上位にランキングされている記事に関しては、ほとんど差はないが、下位になるほど多様な記事が増加することなどがわかった。

本研究では、読者が記事を読んで興味を持ったか否かの情報が反映される点と約半分はアクセスランキングに存在しない記事が含まれている点を考慮し、SNS での注目度を中心に推薦に利用することとする。

4. 教科書の見出しの有用性調査

SNS での注目度を利用したランキングは、あくまで一般ユーザの視点で構成されることになるため、NIE に適したランキングになるとはいえない。また、ランキングを構成する記事が、NIE に適した記事である保障もない。そこで、NIE にとって不要な記事はフィルタリングし、また、教育的に価値のある記事は、上位に順位されるような仕組みが必要となる。

我々の先行研究では、SNS の注目度に加え、子供向けサイトの検索ログを利用したランキング手法[5]を提案した。本稿では、教科書の見出しから抽出した単語がニュース記事に含まれる割合を調査し、教育的知識として利用できる可能性について考察する。

4.1. データ

調査には、2014 年 6 月から 12 月まで産経ニュースサイトと YOMIURI ONLINE から収集した 26,908 件の記事を利用する。教科書は、東京書籍の新編新しい社会 6 下[4]を利用する。教育的知識として、教科書の単元情報(見出し)に含まれる名詞 122 語を利用する。なお、1 文字からなる名詞はノイズになりやすいため、対象から取り除いている。

4.2. 結果と考察

調査の結果、26,908 件中 20,674 件(76.8%)に教育的知識のいずれかの単語が含まれていることを確認した。含まれていた単語の上位 3 件は“日本”“世界”“関係”といった抽象的な語であった。“日本”は 9,026 記事、“世

界”は 3,849 記事、“関係”は 3,748 記事に含まれていた。また、“基本的人権”“核兵器”“環境問題”などの具体的な語を含んでいる記事も多数確認できた。

以上より、約 77%の記事に教育的知識が含まれていることを確認したが、上位には抽象的な語が多い結果となった。抽象的な語は、ノイズになりやすいため、教科書の見出し情報から得られる単語の選別が必要であると考えられる。また、教科書の本文から重要語を抽出し、教育的知識に加えることも可能である。今後は、教科書だけでなく、他からも教育的知識を拡充する方法を検討する。

5. まとめ

本稿では、まず、ニュースサイトにおけるアクセスランキングとソーシャルランキングの傾向分析を行った。分析の結果、アクセスランキングとソーシャルランキングでは、平均して 51.0%の記事が異なることを確認した。また、上位にランキングされている記事に関しては、ほとんど差はないが、下位になるほど多様な記事が増加することも確認した。

次に、教科書の見出しから抽出した単語がニュース記事に含まれる割合を調査し、教育的知識として利用できる可能性について考察した結果、76.8%の記事が見出しに含まれる名詞を含んでいることを確認した。しかし、多数のニュース記事に含まれる語は抽象的な語が多いため、教科書の見出し情報から得られる単語の選別が必要であることがわかった。教育的知識としての効果が期待できる具体的な語を拡充するために、教科書の本文からも重要語の抽出を行う必要がある。

今後の課題としては、教育的知識の拡充と、先行研究に教育的知識を活用する手法を検討する。そして、記事推薦システムを構築し、提案手法の有用性を評価する。

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

- [1] NIE 教育に新聞を、<http://nie.jp/>
- [2] 坪井賢泰, 安藤一秋, “小学生向け NIE を対象とした Web 新聞記事の推薦”, 情報処理学会研究報告, Vol.2013-CE-119, No.18, pp.1-5, 2013.
- [3] H. Kwak, C. Lee, H. Park, and S. Moon, “What is Twitter, a social network or a news media?”, *Proc. of the 19th international conference on World wide web*, pp.591-600, (2010).
- [4] 新編 新しい社会 6 下, 東京書籍, (2012)
- [5] S. Tanaka, and K. Ando, “Web News Ranking for Elementary School Children based on Degree of SNS Users' Attention and Popular Search Queries among Children”, *Proc. of 6th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, pp.475-478, (2015).