

SNS での注目度に基づいた 小学生のための Web ニュースランキング手法

田中 翔也[†]安藤 一秋[‡]香川大学大学院工学研究科[†]香川大学工学部[‡]

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するが、近年では、新聞を購読しない家庭も増えているため、Web ニュースを利用する機会も増えてきた。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。したがって、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].

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るため、読者が記事に興味をもてば、容易に投稿可能である。代表的な SNS として Twitter や Facebook などがあり、特に Twitter ではツイートの 85% がニュース記事に関連するものであるという報告[3]がある。SNS での注目度が高いニュースは、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、SNS での注目度と子供に人気のある検索語を用いた Web ニュース記事のランキング手法について妥当性を評価する。

2. ニュースサイトのアクセスランキングと SNS での注目度に基づくランキングの違い

Web 新聞サイトには、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、読者が記事を読んで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シェアされる可能性は低いと考えられるため、より興味をもたれた記事のランキングが生成できると考えられる。

A Web News Ranking Method based on Degree of Attention of SNS users for Elementary School Children

Shoya TANAKA[†], Kazuaki Ando[‡][†] Graduate School of Engineering, Kagawa University,[‡] Faculty of Engineering, Kagawa University

しかし、これらのランキングは、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。

3. SNS での注目度と子供語を用いたランキング

3.1. 子供語

Yahoo!きっずやキッズ goo などの子供向けポータルサイトでは、検索語のランキングを公開している。このランキングは、子供が興味を持つ言葉のランキングとみなせるため、これらの語を含む記事は、子供が興味を持つ可能性があると考えられる。また、検索語ランキングの順位情報は、重み情報としても利用可能である。

以下、記事をランキングするための重要度計算について説明する。以降、子供向けポータルサイトにおいて人気のある検索語を子供語と呼ぶ。

3.2. 子供語に基づく記事の重要度

記事の重要度は、記事内の単語の重要度と子供語の重みを掛けた値の総和として算出する。単語の重要度は、Okapi BM25[4]と TFIDF で計算し、実験により、いずれかを選択する。

記事 i に含まれる単語を t_j 、その総数を n とするとき、記事の重要度 $S(i)$ は、式(1)と式(2)で計算される。

$$S(i) = \sum_{j=1}^n C(t_j) \cdot \text{BM25}(t_j) \quad (1)$$

$$S(i) = \sum_{j=1}^n C(t_j) \cdot \text{TFIDF}(t_j) \quad (2)$$

ここで、 $C(t_j)$ は子供語の順位情報に基づく重みである。なお、式(2)の TF 値は、記事中の単語総数で正規化した値を利用する。

$C(t_j)$ は、子供語の順位情報を基に、式(3)で計算される。

$$C(t_j) = \frac{p \cdot \log_r(r + \text{rank}_{\max} + 1 - \text{rank}(t_j))}{1} \quad (3)$$

ここで、 p と r は調整用パラメタである。 rank_{\max} は、利用する子供向けポータルサイトにおける子供語の最大順位であり、 $\text{rank}(t_j)$ はそのサイトにおける t_j の順位である。式(3)は、子供語の順位が高い程、重みが大きくなる。もし、 t_j が子供語でない場合、 $C(t_j)$ の値は 1 とする。

3.3. SNS での注目度と子供語に基づく記事の重要度

記事 i に対する SNS での注目度 $SNS(i)$ は、記事に対する Tweet 数と Facebook でのシェア数を基に、各々を正規化した値の和として算出する。

最終的な記事 i の重要度 $Score(i)$ は、子供語に基づく記事の重要度 $S(i)$ と $SNS(i)$ を用いて、以下の式で計算される。

$$Score(i) = SNS(i) \cdot S(i) \quad (4)$$

$$SNS(i) = Twitter(i) + Facebook(i) \quad (5)$$

ここで、 $Twitter(i)$ は記事 i に対する Tweet 数、 $Facebook(i)$ はシェア数をそれぞれ対象期間の最大値で正規化した値を利用する。

最終的には、 $Score(i)$ に基づき各記事をランキングする。

4. 評価

4.1. 概要

式(4)に基づくランキング (Proposed) と式(5)のみを利用したランキング (Baseline) を被験者 (大学生) 24 人に匿名で提示し、NIE に適したランキングを選択してもらい、ランキングの妥当性を評価する。また、選択時には、その理由も記述してもらう。

評価対象の記事 (読売, 産経) は、2014 年 12 月 1 日から 7 日に収集し、NIE に不適切な語 (事件や事故など) を含む記事を人手でフィルタリングした 3,470 件である。また、子供語 (Yahoo!きっずやキッズ goo) も同じ期間のものを利用する。

評価セットを表 1 に示す。期間はランキング生成に利用した記事の期間を示す。記事数による影響、単語の重み付け手法を比較するため、4 つの評価セットを利用する。なお、式(3)のパラメタは、仮に $r = 2$, $p = 2,500$ とする。

表 1. 評価セット

セット	期間	重み付け
1	1 day	TFIDF
2	1 day	BM25
3	1 week	TFIDF
4	1 week	BM25

4.2. 結果と考察

4 つの評価セットに対して、被験者が選択したランキング数 (得票数) を図 1 に示す。図 1 より、全セットに対して Proposed の得票数が Baseline より多く、全体で約 67% (49/73 票) を得た。

提案手法のランキングが選択された理由としては、「子供にわかりやすそうな記事が多い」「子供が興味を持ちやすい記事が選ばれている」「子供向けの記事が多い」といった理由が多かった。一方、選択されなかった理由としては、「エンタメ色が強い」「コメディ、キャラクター要素が多く、堅い記事が少ない」といった理由があった。

総合結果としては、提案手法のランキングが妥当であると評価されたが、本稿で提案した手法は、

NIE 向けのランキングを生成するための要素として、子供語による重み付けしか導入していない。つまり、子供が興味を持つ記事を上位にランキングすることは可能であるが、NIE に適しているという点では、まだ改善が必要である。この点が得票数を伸ばせなかった原因として考えられる。

今後は、教科書の重要語や単元情報、教師からのキーワード指定などの学習・教育的要素を導入し、小学生に興味を持ってもらいたい記事や授業内容に関係のある記事など、より NIE に適した記事をランキングできるよう手法を改善する。

最後に、単語の重要度計算については、記事数が多くなると TFIDF の得票数が減少しているが、BM25 は安定している。今後は、詳細な実験および考察を行い、手法を選択する必要がある。

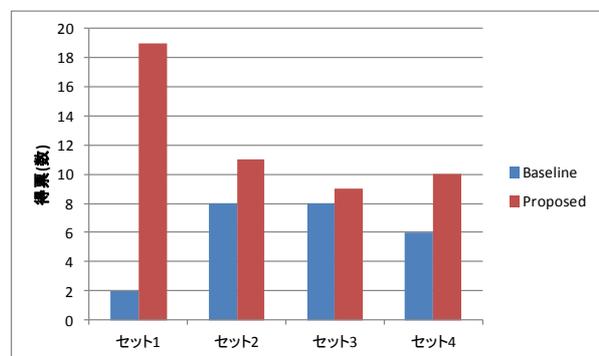


図 1. 評価結果

5. おわりに

本稿では、SNS での注目度と子供に人気のある検索語を用いた Web ニュース記事のランキング手法について、妥当性を確認した。今後の課題として、単語の重み付けの選択の他、教科書の重要語や単元情報などによる重み付けを導入し、より NIE に適した記事を推薦できる手法を実現する。

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] Kwak, H., Lee, C., Park, H., and Moon, S., “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] S. Robertson and S. Walker, “Okapi/Keenbow at TREC-8”, Proc. of the 8th Text Retrieval Conference, 1999.