

小学校教師に対する Web ニュース推薦に向けた NIE ワークシートの分析と NIE 教材に適した記事判定法の検討

Analysis of NIE Worksheets and Article Estimation Method for Web News Recommendation to Elementary School Teachers

関 伸也*¹
Shinya Seki

安藤 一秋*²
Kazuaki Ando

*¹ 香川大学大学院工学研究科
Graduate school of Engineering, Kagawa University

*² 香川大学工学部
Faculty of Engineering, Kagawa University

In elementary schools, NIE (Newspaper in Education), which uses newspaper as teaching materials in classes, is being implemented. However, teachers' loads are increasing by adjusting time in general classes, selecting suitable articles and preparing support materials. The purpose of this study is to construct a Web news recommendation system for elementary school teachers. This paper analyzes articles in NIE worksheets based on categories and readability, and examines an estimation method of Web news suitable for NIE in elementary schools.

1. はじめに

NIE (Newspaper in Education) は、新聞を教材として活用する取り組みのことであり、小学校や中学校を中心に、幅広い教育機関で実施されている[NIE 17]。各地域の NIE 推進協議会は、NIE 実践報告書[報告書 17]を毎年発行している。小学校での実践報告によると、地元や周辺地域に関する記事を用いて、地域学習として NIE を実践している例が多い。また、新聞記事を選ぶ・読む・内容について話し合うなどの活動を通じて、児童の読解力・語彙力・表現力・思考力や社会への興味関心が向上したと述べられている。さらに、物事を多角的にとらえる力・根拠を示しながら自分の考えを述べる力・情報を調べて整理する力の育成に効果があることも確認されている。

小学校での NIE では、各新聞社が発行している紙媒体の新聞や Web ニュースを使用している。しかし、これらの記事は児童を対象として書かれていないため、児童が学習していない漢字や、理解することが困難な表現などが使われており、児童にとって難解な文章であることが多い。そのため、日々発行される膨大な記事の中から、児童自身が自分の興味や課題に適した記事を探し出すことが難しいという実践課題がある。また、教師にとっては、NIE に適した記事を探すことや、記事に関する教材研究を行う必要があることで、教材準備の負担が増加するといった問題がある。したがって、記事の推薦や難解な言葉の言い換えなど、児童に対する支援と、NIE に活用しやすい記事や関連・補足資料の検索・選択など、教師に対する支援が必要になる。そのため、新聞記事に関する読解支援[河村 17]や児童に対する記事推薦[田中 16]などの研究が行われている。また、NIE の補助教材として、NIE ワークシートを Web 公開することで、授業実践を補助している地域紙はあるが、数が少ない。

本研究では、NIE を実践する教師向けの支援に注目し、NIE に活用しやすい Web ニュース記事を地域学習の教材集として推薦するシステムの構築を目的とする。NIE 教材としての価値が高い記事を推薦することで、教師が記事を探す負担を軽減し、NIE を継続的に実践する補助になると考える。

本稿では、NIE ワークシートを分析し、NIE 教材に適した記事を判定する手法について初期検討する。

2. NIE ワークシート

NIE ワークシートとは、NIE を推進するための支援サービスの 1 つである。一部の新聞社は、NIE 実施の補助教材として独自にワークシートを作成し、Web 上で公開している。

NIE ワークシートは、図 1 のように記事の切り抜きを貼り付けて使用するものと、図 2 のように記事が付随しているものに大別できる。また、記事内容に関する読解問題も付随した NIE ワークシートを公開している新聞社もある。



図 1. 貼付型ワークシート*¹ 図 2. 記事付随型ワークシート*²

日本新聞協会に所属している 104 社[日本新聞協会 17]の Web サイトを調査した結果、15 社で NIE ワークシートが公開されていた。また、表 1 に示すように、15 社中 11 社は、新たなワークシートを継続公開している。11 社の内、「朝日新聞社」、「河北新報社」、「神戸新聞社」、「南日本新聞社」の 4 社は、NIE の経験が豊富な教員の意見も踏まえてワークシートに採用する記事を決定しているため、この 4 社の NIE ワークシートに採用された記事は、NIE 教材として質が保証された記事と考えられる。

しかし、地域学習で NIE ワークシートを利用することを考えると、NIE ワークシートを公開している地域紙は 14 社しかない。したがって、既存の NIE ワークシートのみで地域学習としての NIE 教材が充足できない。そこで、様々な地域の Web ニュースから地域学習に適した記事を推薦できれば、この問題を解決できると考える。

連絡先: 関伸也, 香川大学大学院工学研究科, 安藤研究室,
〒761-0396 香川県高松市林町 2217-20,
s14t242@stmail.eng.kagawa-u.ac.jp

*¹<http://www8.kobe-np.co.jp/nie/2014/08/post-249.html>

*²<http://www8.kobe-np.co.jp/nie/2017/11/17/171116nisikie.pdf>

表 1. NIE ワークシートの公開状況

| 新聞社名 | 公開数 | 教師の 協力 | 現在の 更新 | 新聞社名 | 公開数 | 教師の 協力 | 現在の 更新 |
|---------|------|-----------|-----------|---------|-----|-----------|-----------|
| 朝日新聞社 | 1746 | 確認済 | 有 | 中日新聞社 | 48 | 不明 | 有 |
| 神戸新聞社 | 281 | 確認済 | 有 | 岩手日報社 | 33 | 不明 | 有 |
| 南日本新聞社 | 223 | 確認済 | 有 | 佐賀新聞社 | 17 | 不明 | 有 |
| 河北新報社 | 156 | 確認済 | 有 | 山陽新聞社 | 107 | 確認済 | 無 |
| 徳島新聞社 | 7043 | 不明 | 有 | 山梨日日新聞社 | 27 | 不明 | 無 |
| 静岡新聞社 | 560 | 不明 | 有 | 愛媛新聞社 | 15 | 不明 | 無 |
| 大分合同新聞社 | 230 | 不明 | 有 | 沖縄タイムス社 | 15 | 不明 | 無 |
| 琉球新報社 | 120 | 不明 | 有 | | | | |

3. NIE ワークシートの記事分析

NIE 教材に適した記事の判定法を検討するため、ワークシートに選択された記事の特徴を分析する。本稿では、選択記事のカテゴリの傾向と、記事の読解難易度を分析の観点とする。

分析対象は、現在もワークシートを更新している地域紙の中で、記事選択に教員の意見が取り入れられており、ワークシートの公開数が最も多い、神戸新聞社のワークシートの記事とする。

3.1 記事カテゴリに基づく分析

神戸新聞社の NIE ワークシートの記事 181 件(2015 年 7 月～2017 年 11 月公開)のカテゴリを分析し、NIE 教材に適した記事カテゴリについて考察する。

(1) 新聞社カテゴリの傾向分析

神戸新聞社の NIE ワークシートの記事には、カテゴリが付与されていないが、Web ニュースはカテゴリが付与されている。神戸新聞社の Web ニュースは、“全国海外”、“地域”、“兵庫県内”、“連載・特集”のいずれかのカテゴリに分類され、それらのカテゴリ内でさらに 10 種以上の詳細カテゴリに分類されている。そこで、NIE ワークシートの記事が Web ニュースとして公開されている記事のみを最初の分析対象とする。

NIE ワークシートに掲載されている 181 記事に対し、新聞社カテゴリを調査した結果、表 2 に示すように、181 記事中 57 記事のカテゴリが取得できた。そして、57 記事のうち、地元に関する記事と考えられ、記事数が最も多い「兵庫県内」カテゴリの 41 記事について詳細カテゴリを調べた結果、表 3 に示すように、「社会」カテゴリが最も多いことを確認した。

表 2. カテゴリ内訳

| カテゴリ | 記事数 |
|-------|-----|
| 兵庫県内 | 41 |
| 地域 | 11 |
| 全国海外 | 2 |
| 連載・特集 | 3 |
| 計 | 57 |

表 3. 詳細カテゴリ内訳

| 詳細カテゴリ | 記事数 |
|--------|-----|
| 社会 | 26 |
| 社説 | 7 |
| スポーツ | 3 |
| 文化 | 2 |
| 総合 | 2 |
| 経済 | 1 |
| 計 | 41 |

(2) 人手カテゴリの傾向分析

「社会」カテゴリでは粒度が荒く、NIE ワークシートに採用される記事の傾向がつかめない。そこで、「社会」カテゴリが付与されている 26 記事に対し、人手で、より詳細なカテゴリを付与し、傾向を分析する。

「社会」カテゴリの 26 記事に対して、人手でカテゴリを付与した結果、表 4 に示すように、「生物」や「文化」カテゴリが多いことが確認できた。さらに、181 記事すべてに対して、人手でカテゴリを付与した結果、表 5 に示すように、NIE ワークシートの記事には、「生物」や「文化」の人手カテゴリが付与された記事が選ばれていることが確認できた。以上より、地元に関する記事の中でも、生物の生態に加え、祭り・伝統行事などの文化に関する

記事が地域学習としての NIE 教材によく用いられていると考えられる。

表 4. 人手カテゴリ内訳（「社会」記事）

| 人手カテゴリ | 記事数 |
|--------|-----|
| 生物 | 7 |
| 文化 | 6 |
| 農林水産業 | 4 |
| 自然 | 3 |
| 歴史 | 3 |
| 観光 | 1 |
| 経済 | 1 |
| 社会 | 1 |
| 計 | 26 |

表 5. 人手カテゴリと記事概要（上位 10 カテゴリ）

| 人手カテゴリ | 記事数 | 記事概要 |
|--------|-----|-----------------------------|
| 生物 | 29 | 地元で生息している生物などの生態に関する記事 |
| 文化 | 27 | 正月やお盆などの行事や地元の祭り、伝統工芸に関する記事 |
| 科学技術 | 21 | 宇宙開発やAI、医療などに関する記事 |
| 社会 | 13 | 今年の漢字など日本社会全体に関係のある記事 |
| 農林水産業 | 12 | 地元で作られている作物やその出荷状況に関する記事 |
| 歴史 | 12 | はにわ化石の出土など、歴史的発見に関する記事 |
| 自然 | 11 | 景観地や桜などの自然に関する記事 |
| スポーツ | 11 | スポーツ選手の記録樹立などに関する記事 |
| 学生生活 | 10 | 学生の取り組みや生活実態調査に関する記事 |
| 人物 | 10 | スポーツ選手などへのインタビューをまとめた記事 |

3.2 読解難易度に基づく分析

NIE ワークシートの対象学年と、記事の読解難易度の傾向を分析し、NIE 教材に適した記事の判定に、読解難易度が活用できるか考察する。

分析は、前述の分析で利用した 181 記事から社説や連載記事を除いた 146 記事と、神戸新聞社の Web ニュースサイト[神戸新聞社 18]から収集した 673 記事のリーダビリティスコアを比較することで行う。

(1) リーダビリティスコア

記事文章の読解難易度を表す指標として、以下のリーダビリティスコア R[李 16]を用いる。スコアの算出には、形態素解析機 MeCab^{※3}と形態素解析辞書 UniDic^{※4}を用いる。なお、R の値は、難解な文章ほど低くなる。

$$R = (L \times -0.056 + C \times -0.126 + J \times -0.042 + V \times -0.145 + P \times -0.044) + 11.724 \quad (1)$$

L：平均文長（総形態素数÷文数）

C：漢語率（漢語形態素数÷総形態素数×100）

J：和語率（和語形態素数÷総形態素数×100）

V：動詞率（動詞形態素数÷総形態素数×100）

P：助詞率（助詞形態素数÷総形態素数×100）

(2) リーダビリティスコアの傾向分析

神戸新聞社の NIE ワークシートの対象学年は、8 つのパターンが存在する。このうち、記事数が少ないものを除いた 4 パターンについて、リーダビリティスコアを計測する。

計測結果を表 6 に示す。表 6 から、対象学年が高い、つまり、読解力の高い学年を対象としたワークシートの記事ほど、リーダビリティスコアが低くなっていることがわかる。したがって、児童の読解力に適した記事を判定するために、リーダビリティスコアが寄与する可能性があると考えられる。

※3<http://taku910.github.io/mecab/>

※4<https://ja.osdn.net/projects/unidic/>

表 6. 対象学年ごとのリーダビリティスコアの傾向

| 対象学年 | 記事数 | リーダビリティスコア | |
|--------------|-----|------------|---------|
| | | 平均 | 分散 |
| 小学低学年～小学校高学年 | 29 | 2.96565 | 0.23674 |
| 小学高学年～中学 | 38 | 2.62653 | 0.36023 |
| 小学高学年～高校 | 35 | 2.23461 | 0.34533 |
| 中学～高校 | 23 | 1.86765 | 0.20528 |
| Webニュース記事 | 673 | 2.20321 | 0.72885 |

4. 記事判定法の初期検討

NIE 教材に適した記事の判定法の初期検討として、読解難易度に注目し、2 値分類で高い性能を持つ SVM (Support Vector Machine) を用いて、記事の判定実験を行う。

実験には、OCR を用いて記事本文の抽出を行った NIE ワークシートの記事 146 記事と、神戸新聞社の Web ニュースサイトから収集した記事 98 記事に NIE 教材に適している(正例)、NIE 教材に適していない(負例)のラベルを手で付与した計 244 記事(正例 122 記事, 負例 122 記事)を用いる。

10 分割交差検証における、precision, recall, f-measure, accuracy の平均値で判定性能を評価する。本研究では、判定の正確さとして、precision と accuracy を重視して評価する。

4.1 判定に用いる素性

本稿では、以下の素性を組み合わせることで、NIE 教材に適した記事の判定に有効な素性の組み合わせを検討する。なお、各素性は、平均 0, 分散 1 になるように標準化して利用する。

(1) ベース素性

ベース素性には、記事本文の名詞・動詞・形容詞を利用し、素性値には、TF (Term Frequency) を用いる。

(2) 読解難易度素性

記事文章の読解難易度を表す素性として、式(1)によって算出されるリーダビリティスコアを用いる。

(3) 習得漢字率素性

記事本文中の漢字のうち、小学生が習得していると考えられる漢字の割合を素性に用いる。習得状況は、文部科学省の学年別漢字配当表[文部科学省 18]に漢字が存在しているかで判断する。小学校 1 年から小学校 6 年までの 6 学年ごとに割合を算出し、素性とする。

4.2 実験結果

素性の組み合わせを変えて、10 分割交差検証による、記事判定実験を行った結果を表 7 に示す。表 7 の precision と accuracy より、すべての素性を組み合わせた場合の判定性能が最も高いことがわかる。

読解難易度素性に関しては、ベース素性のみと比較した場合、正例の precision が 0.2 ポイント、負例の precision が 0.8 ポイント、accuracy が 0.4 ポイントと、わずかに向上していることから、NIE 教材に適した記事の判定に影響する素性であることが確認できた。

習得漢字率素性に関しては、ベース素性に追加した場合は、記事判定性能が低下しているが、読解難易度素性と組み合わせた場合は、記事判定性能が向上していることがわかる。これは、読解難易度素性のリーダビリティスコアの算出では、記事文章を構成する形態素の品詞や語種といった情報しか用いておらず、文字そのものの難易度は考慮されていない。しかし、習得漢字率素性によって、文字の難易度を示す指標が加わることで、小学生の語彙力を考慮できるようになったことが影響していると考えられる。

今後は、帯[帯 18]などの読解難易度を算出する他の方法とも比較を行い、さらに読解難易度素性を検討する必要がある。文章の複雑さを表す指標として、記事本文の係り受けを解析し、解析木の深さや葉ノードの数などを素性に利用し、妥当性を検証する予定である。さらに、記事のカテゴリを素性に活用することや、地元に関する記事内容を表す素性として、記事本文中の地名の有無や、その地名の粒度などが記事判定に与える影響なども確認する。

表 7. 記事判定の実験結果

| 素性 | 正例 | | | 負例 | | | accuracy |
|-------------|-----------|--------|-----------|-----------|--------|-----------|----------|
| | precision | recall | f-measure | precision | recall | f-measure | |
| (1) | 0.815 | 0.786 | 0.795 | 0.791 | 0.815 | 0.798 | 0.800 |
| (1)+(2) | 0.817 | 0.794 | 0.800 | 0.799 | 0.818 | 0.803 | 0.804 |
| (1)+(3) | 0.809 | 0.801 | 0.801 | 0.796 | 0.798 | 0.793 | 0.800 |
| (1)+(2)+(3) | 0.822 | 0.809 | 0.810 | 0.812 | 0.816 | 0.808 | 0.812 |

5. おわりに

本稿では、NIE 教材に適した記事を推薦する手法を検討するため、NIE ワークシートを分析し、NIE 教材に適した記事を判定する手法の初期検討を行った。

NIE ワークシートの分析から、地元の生物の生態や伝統行事に関する記事が、地域学習の NIE 教材として、よく用いられる記事であることを確認した。また、低学年向けの記事は、リーダビリティスコアが高くなる傾向を確認した。

次に、NIE 教材に適した記事を判定する手法の初期検討として、SVM による記事判定実験を行った。実験の結果、読解難易度素性と習得漢字率素性を用いることで、正例記事の precision が 0.7 ポイント向上することを確認した。

今後は、他の読解難易度を計測する指標や、記事文章の複雑さ、地名情報の有無、地名の粒度などについて、記事判定の素性としての有用性を検証する。そして、NIE 教材に適した記事を推薦する手法を検討する。

参考文献

- [NIE 17] 教育に新聞を (<http://nie.jp>) : 日本新聞協会。(アクセス日: 2017 年 7 月 26 日)
- [報告書 17] NIE 実験報告書 (<http://nie.jp/report/panflet>) : 日本新聞協会。(アクセス日: 2017 年 7 月 26 日)
- [河村 17] 河村宗一郎, 安藤一秋, “小学生を対象とした Web ニュース読解支援システムのための重要語抽出手法の検討”, JSAI2017 大会論文集, 1J1-5, 2017.
- [田中 16] Shoya Tanaka, Kazuaki Ando, “Web News Recommendation for Elementary School Children using Degree of SNS Users’ Attention and Popular Search Queries among Children”, ACIS International Journal of Computer & Information Science, Vol.17, No.1, pp.17-23, 2016.
- [日本新聞協会 17] 会員者一覧 (<http://www.pressnet.or.jp/member>) : 日本新聞協会。(アクセス日: 2017 年 7 月 26 日)
- [李 16] 李在鎬, “日本語教育のための文書難易度に関する研究”, 早稲田日本語教育学, 第 21 号, pp.1-16, 2016.
- [神戸新聞社 18] 神戸新聞 NEXT (<https://www.kobe-np.co.jp/>): 神戸新聞社。(アクセス日: 2018 年 3 月 4 日)
- [文部科学省 18] 学年別漢字配当表 (http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo3/011/siryu/attach/1400158.htm): 文部科学省。(アクセス日: 2018 年 3 月 4 日)
- [帯 18] 日本語テキストの難易度を計測する (<http://kotoba.nuee.nagoya-u.ac.jp/sc/obi3/>): 佐藤・松崎研究室。(アクセス日: 2018 年 3 月 4 日)