

小学校での NIE 支援に向けた 児童が意見を出しやすい新聞記事の特徴調査

小原 大和[†]
Masakazu Kobara

安藤 一秋[‡]
Kazuaki Ando

1. はじめに

Newspaper In Education (NIE) は、新聞を教材として活用する取り組みのことである。2023 年現在、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。小学校での実践報告[1]によると、NIE を継続することで、読解力や社会に対する関心の向上、情報に対する自己判断力などを養えると報告されている。

小学校での NIE では、新聞社が発行している紙媒体の新聞や Web ニュースが利用されている。また、Web 上で NIE ワークシートを公開している新聞社もあり、授業で活用することもできる。NIE ワークシートは、有識者が選択した新聞記事とその記事内容に関する 3、4 問程度の問題で構成される。しかし、NIE ワークシートを提供している新聞社の数や提供頻度が少ないため、授業で利用したいワークシートが存在せずに教師が自作しなければならない状況が発生している。また、日々発行される膨大な記事集合から NIE に適した記事の選択や補足資料の準備など、教材準備の負担に関する実践課題がある。したがって、記事の推薦や重要語の提示、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。

当研究室では、新聞記事に関する読解支援[2]や地域性に基づく記事推薦[3]などの研究を進めてきた。本研究では、NIE を実践する教師に対する支援に注目し、児童が能動的に意見を出しやすい内容のニュース記事を推薦する技術の実現を目指す。本稿では、児童が意見を出しやすい内容のニュース記事の特徴を検討するため、NIE ワークシートに付随する意見を問う問題に注目して分析する。

2. ワークシート分析

2.1 NIE ワークシートの選定

2022 年 4 月時点で、日本新聞協会に所属する新聞社は 98 社ある。そのうち、NIE ワークシートを公開している新聞社は、2022 年 9 月時点で 16 社あった。表 1 に 16 社が公開している NIE ワークシートの数と更新の有無を示す。

16 社のうち、新しいワークシートが提供されている新聞社は 8 社のみであった。そこで、ワークシートの更新が続いており、提供されているワークシートの数も多い、静岡新聞社のワークシートを調査対象とする。

2.2 問題文の分析

当研究室の先行研果[4]によると、静岡新聞社の提供するワークシートには表 2 の 6 つの問題タイプが存在している。

表 1 NIE ワークシート数と更新状況

新聞社名	NIE ワークシート数	更新
静岡新聞社	1,971	有
南日本新聞社	1,218	有
神戸新聞社	711	有
北海道新聞社	518	有
山梨日日新聞社	314	有
大分合同新聞社	177	有
上毛新聞社	126	有
下野新聞社	118	有
琉球新報社	88	有
佐賀新聞社	199	無
中日新聞社	193	無
山陽新聞社	107	無
岩手日報社	51	無
熊本日日新聞社	49	無
愛媛新聞社	15	無
沖縄タイムス社	15	無

表 2 6 つの問題タイプ

問題タイプ	件数
文章で答える問題	369 (44.1%)
単語で答える問題	265 (31.7%)
穴埋め問題	79 (9.4%)
調査問題	72 (8.6%)
抜き出し問題	26 (3.1%)
選択問題	25 (3.0%)

文章で答える問題の詳細を分析した結果、さらに 5 つの問題タイプに分類できることを確認した。表 3 に問題タイプを示す。本稿では、記事に対して読み手の考えや意見を問う問題が付随する記事に注目し、児童が意見を出しやすい内容のニュース記事の特徴を検討する。

表 3 5 つの問題タイプ

問題タイプ	件数
記事からの読み取り問題	292(31.1%)
記事内容の要約	
記事に対する読み手の考えや意見を問う問題	439(47.0%)
見出しに関する問題	135(14.4%)
写真に関する問題	74 (7.9%)

3. 分類実験のためのデータセット構築

ワークシートの分析より、記事に対して読み手の考えや意見を問う問題が付随する記事の存在を確認した。そこで、意見の出しやすい記事は意見を問う問題が付随する記事であると仮定し、このタイプの問題が付随する記事を正例としてデータセットを構築する。

データセットを構築するため、静岡新聞が提供する 2,642 件 (2022 年 12 月 1 日時点) の NIE ワークシートのうち、940 件を利用する。本稿では、ワークシートに付随す

[†] 香川大学大学院創発科学研究科 Graduate School of Science for Creative Emergence, Kagawa University

[‡] 香川大学創造工学部 Faculty of Engineering and Design, Kagawa University

る問題のうち、以下の 2 つの条件を満たす場合、意見を問う問題と仮定する。

- ・ 解答方法が長文
- ・ 解答例が 1 つでない

人手で 940 件のワークシートを確認した結果、439 件 (47.0%) のワークシートに意見を問う問題が付随していることを確認した。静岡新聞の NIE ワークシートは、PDF ファイルで公開されているため、記事本文のテキストデータを取得するにはコストがかかる。そこで、見出し情報と時間情報を用いて Web 検索することで、本文のテキストデータを Web ニュース記事から収集することにした。

最終的に、Web ニュース記事より構築したデータセットを表 4 に示す。負例については、意見を問う問題が付随していない NIE ワークシートの記事のうち、本文データが収集できた記事で作成した。

表 4 データセットの内訳

	正例	負例	合計
Web ニュース	336	374	710

4. 簡易実験

構築したデータセットを用いて、意見を問う記事がどの程度分類できるかを確認するため、二値分類タスクとして簡易実験を実施する。利用する分類モデルは、ロジスティック回帰 (lr) と Support Vector Machine (svm) である。

記事ベクトルとして、単語の出現有無に基づく Bag-of-Words (bow) と tf-idf を重みとする Bag-of-Words (tf-idf)、さらに、Wikipedia を用いて skip-gram で学習した 200 次元の word2vec (w2v) を使用する。実験設定として、学習データの正例と負例が 1 : 1 となるように 252 件ずつにダウンサンプリングする。なお、本文ベクトルには、名詞、動詞、形容詞のみを用いる。

学習データとテストデータを 3 : 1 で分割し、学習データに対して、5 分割交差検証で評価し、汎化性能が高いモデルを選択する。そして、テストデータに対する分類性能を precision と recall, f1 で評価する。

4.1 実験

実験結果を表 5 に示す。最良の precision は、lr_w2v モデルの 0.570 であった。また、どのモデルにおいても f1 が 0.6 を下回る結果となった。そこで、問題タイプの視点からデータセットを再構築して実験することにした。

表 5 実験結果

モデル	precision	recall	f1
lr_bow	0.494	0.488	0.491
lr_tf-idf	0.513	0.488	0.500
lr_w2v	0.570	0.583	0.577
svm_bow	0.458	0.583	0.513
svm_tf-idf	0.542	0.536	0.539
svm_w2v	0.533	0.476	0.503

4.2 データセットの再構築

データセット内の記事を再確認した結果、意見を問わないと判断した記事 (負例) の中に、見出しや写真に関する

問題が付随する記事が含まれていた。しかし、このような問題は、意見を問う問題と問わない問題にかかわらず全ての記事において出題できるといえる。そこで、データセットの負例からこれらを除外することで、データセットを再構築した。そして、再構築したデータセットを用いて再実験した。

再実験の結果を表 6 に示す。表 6 より、ほとんどのモデルの分類性能が向上した。最良の precision は、lr_tf-idf モデルの 0.635 となり、データセットの再構築前と同モデルと比較して 12.2 ポイント増加した。しかし、SVM を用いたモデルの recall が precision より 13 から 40 ポイント高く、ほとんどの記事を正例として予測しているといえる。そのため、NIE ワークシートに採用された記事は、すべて意見を問いやし記事といえる可能性もある。今後は、分類に有用な素性とモデルの検討に加え、意見を問いやし記事の特徴について検討を継続する。

表 6 再実験結果

モデル	precision	recall	f1
lr_bow	0.592	0.525	0.556
lr_tf-idf	0.635	0.500	0.559
lr_w2v	0.613	0.575	0.594
svm_bow	0.581	0.988	0.732
svm_tf-idf	0.540	0.763	0.632
svm_w2v	0.629	0.763	0.689

5. おわりに

本稿では、まず、NIE ワークシートに採用される代表的な問題タイプを確認し、記事に対する読み手の考えや意見を問う問題が存在することを確認した。そして、意見を問う問題が付随する記事を意見の出しやすい記事であると仮定してデータセットを構築し、簡易的な分類実験を実施した。その結果、lr_tf-idf モデルの precision が 0.635 で最良となることを確認した。

今後は、判定手法の検討を続けるとともに、ワークシートに選出される記事は児童が意見の出しやすい記事として扱える可能性についても調査する。また、NIE 実践報告書や学習指導要領より、ディスカッション・ディベートに使用されるテーマに特徴がある可能性があるため、異なる視点での判定手法も検討する。

謝辞

本研究の一部は、JSPS 科研費 22K12294 の助成を受けて実施した。

参考文献

- [1] NIE の学習効果を調べるアンケート結果概要, <https://nie.jp/research/survey/>, 2023 年 6 月 13 日確認
- [2] 河村他, “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.
- [3] S. Seki, et al., “A Method for Determining Web News Suitable as Teaching Materials of Regional Study in Elementary Schools”, IEEJ Transactions on Electronics, Information and Systems, 140(8), pp.964-970, 2020.
- [4] 國方他, “NIE ワークシートの自動生成に向けた問題文の分析”, FIT2022 論文集, pp.281-282, 2022.