

令和 5 年 6 月 9 日現在

機関番号：16201

研究種目：基盤研究(C)（一般）

研究期間：2019～2022

課題番号：19K12271

研究課題名（和文）Webニュースと補助教材，課題の精選を支援する小学校教師のためのNIE支援の実現

研究課題名（英文）NIE Support for Elementary School Teachers: Supporting Selection of Web News, Supplementary Materials, and Assignments

研究代表者

安藤 一秋（Ando, Kazuaki）

香川大学・創造工学部・教授

研究者番号：60325321

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：新聞を教育に活用する教育（Newspaper in Education: NIE）が初等教育機関で実施されている。しかし、NIEを実践する小学校教師は、通常の授業準備に加え、教師自身が授業に適した記事を選定したり、難しい表現を言い換えたり、記事内容を補助する教材を探したりするなど、NIEにかかわる教材準備に対する負担が大きい。そこで本研究では、NIEを実践する小学校教師の教材準備に関する負担を軽減するため、NIEに適したニュース記事を推薦するシステムと、教師が選択した記事の内容を補足する画像コンテンツを推薦するシステムを提案する。評価実験結果から提案システムの有効性を確認した。

研究成果の学術的意義や社会的意義

NIEに関する教材準備の支援環境を構築することで、小学校教師の教材準備に関わる負担を軽減できるだけでなく、未知の教材コンテンツに出会える機会も増やせるため、新しいNIE実践の道を切り開くことができる。また、提案するシステムは、中学、高校の教師でも活用できる。さらに、提案する要素技術は汎用性があるため、児童の記事選択や読解支援、調べ学習支援などのNIE支援への応用だけでなく、NIE以外の教材準備、一般ユーザへのサービスなどに幅広く展開できるといえる。

研究成果の概要（英文）：Newspaper in Education (NIE) is an educational approach that utilizes newspapers as teaching materials. NIE is effective in elementary school education, however, it places a significant burden on teachers in terms of preparing materials specific to NIE. To address this issue, we proposed a system that recommends news articles suitable for NIE to elementary school teachers. Additionally, we proposed a system for recommending images that complement the content of news articles. Based on the experimental results, we have confirmed the effectiveness of the proposed systems.

研究分野：自然言語処理，情報検索，教育応用

キーワード：Newspaper In Education Webニュース 小学校教師 NIE支援 補助教材 画像コンテンツ

1. 研究開始当初の背景

新聞を教育に活用する教育 (Newspaper in Education: NIE) が初等教育機関で実施されている。日本新聞協会が実施した調査と各地域の NIE 推進協議会の報告書によると、NIE を実践することで、新聞に対する読解習慣が身に付き、読解力の向上やコミュニケーション能力の育成などにつながると報告されている。また、平成 23 年度以降は、学習指導要領にも「教材としての新聞の活用」が組み込まれるだけでなく、新しい教科書にも新聞を題材にした学習内容が増強され、NIE への注目は高い。

しかし、一般の新聞記事を小学校教育で利用する場合、児童にとって読めない漢字や意味のわからない表現があるため、記事を探せない、読んでも理解できない等の実践課題が存在する。また、多忙な教師にとって、通常の授業準備に加え、教師自身が NIE 授業に適した記事を選定したり、難しい表現を言い換えたり、記事内容を補助する教材を探したりするなど、NIE にかかわる教材準備に対し、さらに時間と労力を要するという実践課題がある。

2. 研究の目的

研究代表者は、これまで、平成 25-27 年度科学研究費基盤研究(C)などの支援を受け、児童が記事を読めない、探せない問題に注目して、興味ある記事の推薦や読解支援など児童に対する支援に取り組んできた。本研究では、児童ではなく、NIE を実践する小学校教師の教材準備に関する課題に注目し、NIE に適したニュース記事の推薦と、記事内容を補足する画像コンテンツ(図、写真、グラフなど)の推薦により、小学校教育における NIE 教材の準備を支援するシステムの構築を目指す。本研究により、教材準備に係わる教師の負担を軽減できるだけでなく、未知の教材コンテンツに出会える機会も増やせるため、新しい NIE 実践の道を切り開くことができる。

3. 研究の方法

NIE を実践する教師の教材準備を支援するため、当該研究期間において、主に以下の研究に取り組んだ。

3. 1 NIE に適したニュース記事の判定と推薦システムの構築

推薦システムの構築にあたり、まずは NIE に適したニュース記事の判定法を検討する必要がある。そこで、本研究では、新聞社が発行する NIE ワークシートに注目する。NIE ワークシートには、新聞記事と記事内容に関する問題が 3 問程度記載されている。ワークシートの作成には、NIE の経験が豊富な教師が携わっていることもあり、採用された記事は、NIE 教材として質の高い記事といえる。研究開始当初に、日本新聞協会に所属する所属する新聞社の Web サイトを確認した結果、14 社がワークシートを公開していた。そこで、14 社の NIE ワークシートを収集し、採用された記事の特徴(カテゴリ、難易度、地域との関連性など)を分析する。その結果をもとに、NIE に適した記事を推薦するための素性を検討し、モデルを構築・評価する。その後、最良モデルを用いて推薦システムを構築する。

3. 2 記事内容を補足する画像コンテンツ推薦システムの構築

ニュース記事の内容に関連する画像コンテンツは、Web 上に多数存在していると考えられる。そこで、本研究では、記事内容や重要語などを補足する視点として、以下の 2 点に注目する。

(1) ニュース記事との関連性

(2) ニュース記事の内容を補足・説明する内容であるか(解説性)

そして、教師が選択した記事に対し、記事内容を補足する画像コンテンツを、画像の周辺テキストと記事内容との関連性、解説性に基づいてスコアリングし、グラフ、図、表、その他の 4 タイプに分類してランキング提示するシステムを構築する。

3. 3 ニュース記事の言い換えに関する検討

教師の記事選択の視点から「重要だけど表現等が難しいから選ばない」を排除するために、一般のニュース記事を平易に言い換える技術の実現も必要となる。そこで、これまでに構築していた言い換えデータセットを拡充するため、NHK ニュースおよび NEWS WEB EASY サイトから新規に収集した記事対を対象に、それぞれの文対を求めるための文間類似度計算法について検討する。

3. 4 SNS ユーザの注目度の活用に向けたユーザ居住地の推定に関する検討

教師への新たな記事推薦の観点として、居住地に関連するニュース記事の SNS ユーザの注目度について調査する。ニュース記事について言及する SNS ユーザの居住地を推定する手法を検討する。特に、日本人ユーザの多くは、自身のプロフィールに居住地を明記せず、ツイートにも位置情報をほとんど付与せず、地域を特定する語をあまり利用しない傾向がある。したがって、発信者の位置情報を推定するためには、Twitter 内外の情報を活用する必要がある。そこで、本研

究では、Twitter 外の情報として、天気情報と地震情報に注目し、ツイート本文とこれらの情報を用いて、深層学習モデルによる居住地を推定する手法について検討する。

4. 研究成果

4. 1 NIE に適したニュース記事の推薦システムの構築に関する成果

まず、14 社のワークシートに採用された記事の特徴（カテゴリ、難易度、地域との関連性など）を分析し、NIE に適した記事を推薦するための特徴量として、以下の素性を検討した。

- (1) Juman++の解析結果として得られるカテゴリと日本語語彙体系の意味カテゴリの記事中での出現割合 (C)
- (2) Web 上のまとめサイトから収集した都道府県ごとの「特産品」、「祭事」、「文化財」、「工芸品」に関する特徴語の記事内頻度 (W)
- (3) 記事の読みやすさを示すリーダビリティスコア (R)

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

- L : 平均文長 (総形態素数 ÷ 文数)
- C : 漢語率 (形態素数 ÷ 総形態素数 × 100)
- J : 和語率 (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 (形態素数 ÷ 総形態素数 × 100)

※Y. Hasebe, et al. “Introducing a readability evaluation system for Japanese language education”, CASTEL/J 2015. より引用

- (4) 学年別漢字配当表 (文部科学省) を用いた学年ごとの学習済み漢字の出現割合
- (5) 日本語教育語彙表に基づく、語彙レベル漢字毎の出現割合
- (6) 文書ベクトルとして、単語の出現頻度に基づく **Bog of Words** ベクトル (BoW) または **Word2Vec** の平均ベクトル (W2V)

以上の素性を用いて、**Support Vector Machine (SVM)**、ロジスティック回帰、ランダムフォレスト、k-近傍法、ナイーブベイズ、AdaBoost、決定木を用いて、判定性能を評価した。神戸新聞社、徳島新聞社、中日新聞社、静岡新聞社、南日本新聞社の 5 社のワークシートに付随している記事を正例 (計 428 件)、各新聞社の Web サイトから自動収集した記事群から人手で NIE の地域学習に適さないと判断した記事を負例 (計 428 件) からなるデータセットを構築して評価に利用した。

まず、10 分割交差検証により、7 つのモデルの性能を評価した結果、SVM を採用したモデルの F 値および適合率がそれぞれ 0.925 と 0.944 と最良値を得た。その後、SVM を用いて、6 素性の組み合わせを 10 分割交差検証で評価した。

表 1 に適合率の上位 10 件の結果を示す。評価の結果、Word2Vec と学習済み漢字割合 (K) を用いた場合、適合率が最良となることを確認した。また、適合率上位の結果のほとんどには、学習済み漢字割合 (K) が含まれていることから、この素性が特に有効であるといえる。

表 1. SVM による判定結果

素性	適合率	再現率	F 値
W2V+K	0.950	0.954	0.952
W2V+RKGW	0.949	0.943	0.946
BoW+KCW	0.949	0.921	0.934
W2V+KW	0.949	0.945	0.947
W2V+RG	0.947	0.947	0.947
W2V+KGW	0.947	0.945	0.946
W2V+RK	0.947	0.950	0.948
BoW+RKC	0.947	0.924	0.934
BoW+RKCW	0.947	0.921	0.933
BoW+KCGW	0.947	0.921	0.933
BoW+RKCGW	0.947	0.921	0.933
W2V	0.940	0.942	0.941
BoW	0.927	0.901	0.913

次に、データセットには含まれない未知記事を対象に、最良モデルの性能を評価した。ニュース記事が収集できる 30 社の新聞社サイトから約 11 万記事を収集し、モデルが正例と判定した上位の結果の内、判定値の高い上位 1,000 記事を主観評価した。その結果、0.978 の適合率が得られ、提案手法の有効性を確認した。

最良手法を用いて、図 1 に示す記事推薦システムを構築した。提案システムは、記事収集機能、記事判定機能、全文検索機能、カテゴリ絞り込み機能をもつ。教師は、提案手法により算出したスコアでランキングされた記事を地域別やカテゴリ別に閲覧できる。また、任意のキーワー

ドで全文検索できる。提案システムを利用することで、NIE に適した記事のみがおすすめ順で提示されるため、記事選別コストを軽減できる。

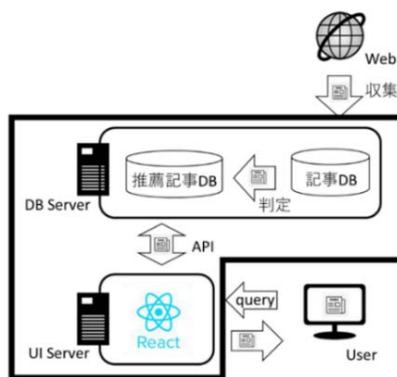


図 1. ニュース記事推薦システムの構成図

4. 2 記事内容を補足する画像コンテンツ推薦システムの構築に関する成果

提案するシステムは、教師が選択した記事に対して、重要語を抽出し、画像検索クエリを生成する。そして、生成したクエリを使って画像を検索し、記事との関連性と解説性に基づいて画像をリランキングして提示する。図 2 に記事内容を補足する画像コンテンツ推薦システムの構成図を示す。

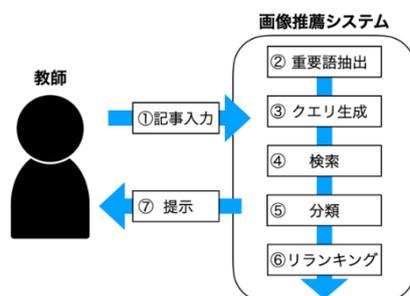


図 2. 記事内容を補足する画像コンテンツ推薦システムの構成図

教師が補足画像を必要とする記事を入力すると、Okapi BM25 を用いて重要語を抽出する。そして、記事本文や Wikipedia 記事から得られる共起語や補足語などを用いて構築した規則（例えば、「重要語+の+サ変」や「重要語+の+共起語」）で、複数のクエリを生成する。その後、Web 画像検索 API を用いて、生成したクエリに基づく画像群を収集する。そして、収集した画像をグラフ、図、表、その他の 4 つのタイプに自動分類した後、関連性と解説性に基づきリランキングしてユーザに提示する。

関連性については、TF-IDF を重みとする Bag-of-Words ベクトル、画像周辺テキストと記事の共通語の出現回数、画像周辺テキストにおけるニュース記事の重要語の出現回数、画像周辺テキストにおける重要語に対する属性、動作を表す表現の出現回数などを基にスコアリングする。

解説性については、画像周辺テキストが用語説明文、比較文、データ解説文のいずれかの特徴をもつ場合、解説性ありと判定する。解説性の有無は、二値分類問題として、Support Vector Machine を用いて判定し、その予測値をスコアに利用する。Bag-of-Words ベクトルまたは word2vec の平均ベクトルをベース素性に利用する。また、追加素性として、3 タイプの文（用語説明文、比較文、データ解説文）で頻出する表現を日本語 WordNet で拡張した辞書として定義し、その素性値には、テキスト内での出現頻度を利用する。

100 画像を用いて、関連性と解説性を評価した結果、関連性については TF-IDF を利用したスコアリングが最良となり、解説性については word2vec に追加素性を利用した手法が最良となった。そこで、推薦システムでは、この 2 つのスコアを用いて、画像をリランキングする。

提案システムの有効性を確認するため、Google 画像検索と提案システムを 6 名の被験者に利用してもらい、アンケートにより評価した。なお、実験結果については、現在、論文を執筆中であるため、割愛する。

4. 3 ニュース記事の言い換えに関する検討の成果

NHK NEWS WEB と、その記事を平易に言い換えた NHK NEWS WEB EASY から、特性の異なる文間類似度計算法（Additive Embeddings, Average Alignment, Maximum Alignment, Hungarian Alignment, Word Mover's Distance, Word Rotator's Distance, sentence-BERT）を用いて難解文と平易

文対を対応付けて、複数の平易化用パラレルコーパスを構築した。そして、それぞれのパラレルコーパスを学習データに利用して、複数の平易化換言モデルを構築した。その後、各モデルの性能を評価することで、テキスト平易化における文間類似度計算法とパラレルコーパスの性質を分析した。換言割合・文長・単語数・文節数などに着目してコーパス特性を分析した結果、Average Alignmentに基づく文間類似度計算法を用いることにより、多様な言い換え対を含むパラレルコーパスを構築できることを確認した。

4. 4 SNS ユーザの注目度の活用に向けたユーザ居住地の推定に関する検討の成果

ツイート内部の情報と天気、地震情報に基づく深層学習モデルを検討した。図2にモデル図を示す。モデルの詳細説明は、主な発表論文に記載の論文に譲る。

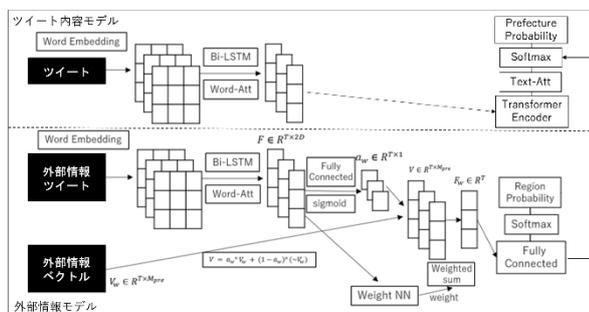


図3. ユーザ居住地の推定モデルの構成図

プロフィールに自身の居住地を明記している 204,965 ユーザのうち、学習データを 184,965 ユーザ、検証データ、テストデータを、それぞれ 10,000 ユーザに利用して、検討したモデルの性能を評価した。

47 都道府県別予測結果の Precision, Recall, F 値を表2に示す。表2より、ツイート内容のみを用いる場合と比べ、ツイート内容に加えて天気情報および地震情報を追加することで都道府県別予測結果の F 値が約 2.3 ポイント向上していることがわかる。以上の結果から、ツイート本文に地域を特定する語を利用しないユーザに対しても、天気や地震情報を利用することで、居住地の推定性能を向上することができるといえる。

今後は、このモデルを利用して、居住地に関連するニュース記事の SNS ユーザの注目度について調査する。そして、注目度を利用したニュース記事の推薦手法について検討する。

表2. 都道府県別予測結果

	Precision	Recall	F 値
ツイート内容のみ	0.7245	0.6929	0.7046
+天気情報	0.7730	0.6899	0.7244
+天気・地震情報	0.7835	0.6873	0.7277

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 0件）

1. 著者名 Seki Shinya, Ando Kazuaki	4. 巻 140
2. 論文標題 A Method for Determining Web News Suitable as Teaching Materials of Regional Study in Elementary Schools	5. 発行年 2020年
3. 雑誌名 IEEJ Transactions on Electronics, Information and Systems	6. 最初と最後の頁 964 ~ 970
掲載論文のDOI（デジタルオブジェクト識別子） 10.1541/ieejieiss.140.964	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計12件（うち招待講演 0件/うち国際学会 4件）

1. 発表者名 西山 朋貴, 安藤 一秋
2. 発表標題 子供向け日本語換言コーパス構築のための文間類似度計算法の評価
3. 学会等名 2021年度人工知能学会全国大会
4. 発表年 2021年

1. 発表者名 國方 海秀, 安藤 一秋
2. 発表標題 Webニュースからの課題作成支援に向けたNIEワークシートの分析
3. 学会等名 2021電気関係学会四国支部連合大会
4. 発表年 2021年

1. 発表者名 Taiki Oguri, Kazuaki Ando
2. 発表標題 A Prototype System for Recommending Images Supplementing Web News for NIE in Elementary Schools in Japan
3. 学会等名 IEEE TALE2020: An International Conference on Engineering, Technology, and Education (国際学会)
4. 発表年 2020年

1. 発表者名 Shinya Seki, Kazuaki Ando,
2. 発表標題 Examination of a Method for Determining Web News Suitable as Teaching Materials of NIE toward News Recommendation for Elementary School Teachers
3. 学会等名 5th International Conference on Electronics and Software Science (国際学会)
4. 発表年 2019年

1. 発表者名 Ryuya Ikedi, Kazuaki Ando
2. 発表標題 Extraction of Food Product and Shop Names from Blog Articles using Named Entity Recognition
3. 学会等名 16th International Conference of the Pacific Association for Computational Linguistics (国際学会)
4. 発表年 2019年

1. 発表者名 関伸也, 安藤一秋
2. 発表標題 小学校でのNIE教材に適したWebニュース記事判定手法
3. 学会等名 電子情報通信学会 教育工学研究会
4. 発表年 2019年

1. 発表者名 池田流弥, 安藤一秋
2. 発表標題 固有表現抽出によるブログテキストからの品名・店名抽出
3. 学会等名 情報処理学会 自然言語処理学研究会
4. 発表年 2019年

1. 発表者名 小栗太樹, 安藤一秋
2. 発表標題 小学校におけるNIEのためのWebニュースを補足する画像のスコアリング手法の検討
3. 学会等名 第18回情報科学技術フォーラム
4. 発表年 2019年

1. 発表者名 関伸也, 安藤一秋
2. 発表標題 小学校での地域学習NIEに適したWebニュース記事推薦システム
3. 学会等名 情報処理学会第82回全国大会
4. 発表年 2020年

1. 発表者名 小栗太樹, 安藤一秋
2. 発表標題 Webニュースを補足する画像コンテンツの推薦に向けた画像周辺テキストの解説性に基づくスコアリング手法
3. 学会等名 情報処理学会第82回全国大会
4. 発表年 2020年

1. 発表者名 Masahiro Matsumoto, Kazuaki Ando
2. 発表標題 A Deep Learning Model of Estimating User ' s Place of Residence Using Tweets and Weather Information
3. 学会等名 2022 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (国際学会)
4. 発表年 2022年

1. 発表者名 松本 真拓, 安藤 一秋
2. 発表標題 ツイートの内容と天気・地震情報を用いたTwitter ユーザの居住地推定
3. 学会等名 情報処理学会第85回全国大会
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	溝淵 昭二 (Mizobuchi Shoji) (30340756)	近畿大学・情報学部・准教授 (34419)	
研究協力者	関 伸也 (Seki Shinya)	香川大学・工学研究科・学生 (16201)	
研究協力者	小栗 太樹 (Oguri Taiki)	香川大学・工学研究科・学生 (16201)	
研究協力者	西山 朋貴 (Nishiyama Tomoki)	香川大学・工学研究科・学生 (16201)	
研究協力者	國方 海秀 (Kunikata Kaisyu)	香川大学・創発科学研究科・学生 (16201)	

6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	池田 流弥 (Ikeda Ryuya)	香川大学・工学研究科・学生 (16201)	
研究協力者	松本 真拓 (Matsumoto Masahiro)	香川大学・工学研究科・学生 (16201)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関