Vol.2013-CE-119 No.18
2013/3/16

小学生向け NIE を対象とした Web 新聞記事の推薦

坪井 賢泰[†1]　　安藤 一秋[†2]

本稿では，Web 上の新聞記事を対象に，小学生にとって不適切な記事をフィルタリングする手法と，教師が NIE（Newspaper In Education）で利用することができる記事を推薦する手法について提案する．

Recommendation of Web Newspaper Articles for NIE in Elementary school

YASUHIRO TUBOI[†1]　KAZUAKI ANDO[†2]

This paper proposes a method for filtering Web newspaper articles for elementary school students and a method for recommending newspaper articles which can be used for NIE (Newspaper In Education) to elementary school teachers.

1. はじめに

近年，小学校では，新聞を教材に活用する教育（NIE：Newspaper in Education）が実施されている[1]．しかし，新聞には小学生が読めない漢字や難解な表現が使われているため，小学生が記事を閲読することは難しい．日々，膨大な数の記事が公開されるため，小学生が自分の興味に合う記事を選択することも容易ではない. また, 一般新聞には，犯罪や事故など，道徳・論理教育的に相応しくないと考えられる記事も多数含まれている．これらの問題は，小学生にとって記事の内容理解を困難にさせるだけでなく，記事選択も困難にしている．

NIE を実践する教師は，膨大な数の一般記事から児童個人に読んでもらいたい記事や授業と関連する記事などを選出する必要がある．しかし，授業内容と関連する記事の数はあまり多くないため，膨大な記事群から授業で利用できる記事を探すことは多大な時間と労力を要する．

以上のことから，小学生が単独で新聞記事を読む場合に，不適切となる記事のフィルタリングや，検索力・語彙力をカバーできる記事の検索支援が重要となる．また，教師の負担になっている記事の選出の自動化および推薦は，NIE の実践に有用である．

そこで本研究では，NIE での記事選出を支援するシステムの構築を目的とする．本稿では，Web 上の新聞記事から小学生が単独で閲読する場合，不適切となりうる記事をキーワードおよび SVM（Support Vector Machine）[2]を用いてフィルタリングする手法について提案する．また，教科書の項の見出しと本文を用いて，内容が類似する新聞記事を教師に推薦する手法を提案する．以下，2．では，NIE で利用する新聞の特徴について述べ，3．では，記事選出支援システムの概要を説明する．4．で不適切記事のフィルタリング機能，5．で記事推薦機能を提案する．6.で提案手法の評価を行い，7.でまとめる．

†1 香川大学大学院工学研究科
Kagawa University, Graduate School of Engineering
†2 香川大学工学部
Kagawa University, Faculty of Engineering

2. NIE

NIE とは, 新聞を教育に活用する取り組みのことである．NIE では，新聞の閲読習慣を身につけることで，社会への関心や読解力，メディアリテラシーなど現代社会で必要とされている様々な能力を向上させることを目的としている．

実践例としては，

（1） 読んだ記事の内容について意見交換を行う．
（2） 新聞に載っている知らない単語や語句を調べる．
（3） 興味のある記事を選び，記事について感想を書く．
（4） 実際の新聞の構造を参考に生徒が新聞づくりを行う．

などが挙げられる．

NIE の実践により，「新聞に興味・関心を持つようになった」，「記事について友人・家族と話すようになった」などの効果が得られている[3].

2.1 NIE で利用される新聞

新聞の発行形態は，紙版と Web 版に分けられる．現在の NIE では，紙版の一般新聞を中心に利用している．紙媒体の一般新聞は，一面記事といった場所（ページ）や見出しに書かれる文字の大きさなどにより，ニュースの重要度を認識できる．また，ページ毎に分野を分けて記事を掲載しているため，各分野の記事を俯瞰しやすい．しかし，小学生にとって不適切な記事の除外や各種目的に応じた記事を探す場合，すべての記事の見出し（あるいは内容）に目を通す必要があり，NIE を実践する教師・児童にとって大きな負担となる．

一方，Web 新聞は，最新記事をリアルタイムに得ること

ができる．また，全国紙だけでなく地方紙も無料で閲覧できるため，多くの記事を読むことができる．また，検索機能も備わっているため，目的の記事を検索することが可能である．しかし，紙版のすべての記事が掲載されているとは限らない．また，児童は，語彙力・検索力に差があるため，目的の記事を検索することが容易でない．さらに，広告などの不要な情報が付随しており，クリックすることで不適切なページに誘導されるなどの問題がある．

以上より，紙版に加え，Web 版の新聞を活用することで，NIE の効果をより高めることができるといえるが，Web 版のメリットを活かすためには，不適切な記事・情報の排除や検索・選択支援が必要となる．

2.2 Web 新聞に含まれる不適切記事の割合と特徴調査

Web 新聞に不適切な記事が含まれる割合を把握するため，Web 新聞で扱われている記事の内容を調査する．本調査では，Web 新聞記事を不適切記事と一般的な記事の 2 種類に分類する．不適切記事は，犯罪や事故など倫理教育を終えていない児童に単独では読ませたくない内容の記事，一般的な記事は，それ以外の記事である．調査対象は，毎日 jp で公開されている 1,000 記事とする．

図 1 に調査結果を示す．調査の結果，Web 新聞として公開されている記事の内 30.0%は，児童が単独で記事を閲読する場合，フィルタリングする必要があるといえる．

次に，不適切記事をフィルタリングするための事前調査として，不適切記事に含まれる特徴語を調査する．

調査の結果，不適切記事には，殺人や窃盗など刑法に関連する用語が多く含まれることがわかった．そこで，刑法で規定される罪名を含む割合を調査した．

図 2 に調査結果を示す．図 2 より，不適切記事の 48.0%には，刑法で規定されている罪名が含まれていた．したがって，罪名を利用すれば不適切記事の半数はフィルタリングできる可能性がある．また，罪名を含まない記事の中には，罪名ではなりが，いじめや自殺などが多く含まれていることがわかった．

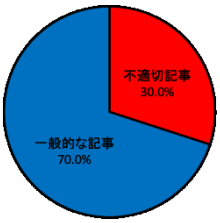

図 1　Web 新聞の分類結果

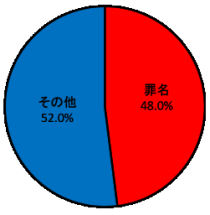

図 2　不適切記事に含まれる単語の特徴

2.3 Web 新聞フィルタリングの必要性

以上の調査により，殺人，傷害，いじめなど，児童に単独で閲読させるには不適切な記事が Web 新聞には複数含まれていることが分かった．これらの記事は，保護者が子供に読ませたくない記事という側面だけでなく，教師にとって授業で扱いづらい内容の記事ともいえる．また，NIE では，教師が選んだ記事を児童に提供する形式の授業だけでなく，児童自らが記事を選択し，その記事中の分からない語や意味を調べる形式の授業も行われている．この様な形式の授業の場合，教師が予め不適切記事を除外した記事集合を作成するなど事前準備に時間を要する．

以上より，NIE を実践する上で不適切記事のフィルタリングは有用であると考えられる．

3. 記事選出支援システムの概要

2.2 での調査により，Web 新聞には不適切記事が全体の 30.0%存在することが分かった．また，小学生は検索経験や語威力，学力などの個人差が大きいため，検索能力にも差がある[4]．また，NIE を実施している教師も，授業で利用したい記事の選出に多大な労力を要する[3]．

以上を考慮して，本研究では，以下の機能を有する記事選出支援システムを提案する．

(1)　カテゴリ検索機能
(2)　キーワード検索に対するサジェスト機能
(3)　記事推薦機能
(4)　不適切記事のフィルタリング機能

本システムの構成を図 3 に示す．

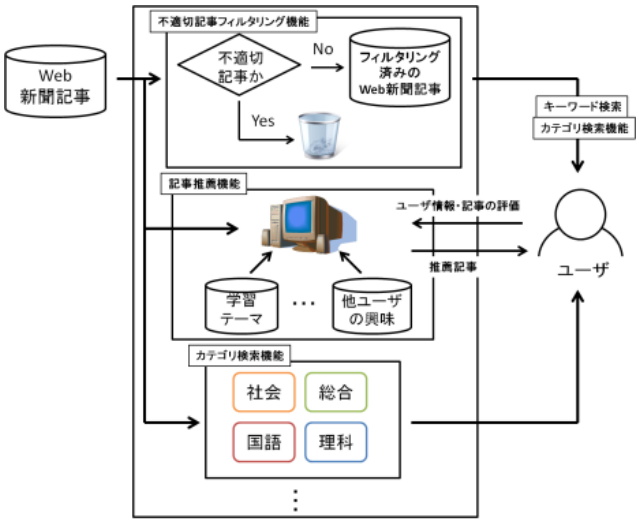

図 3　記事選出支援システム

(1) カテゴリ検索機能

検索力に乏しい児童にとっては，カテゴリから記事を選択することで対応する．Web 新聞記事は，各新聞社によっ

てカテゴリに分類されているが，NIE により適したカテゴリに分類することで，利便性を高める．カテゴリは，読売新聞社が提供している「よみうり博士のアイデアノート」[5] の 6 年生向けページで利用されている社会，総合，国語，理科の 4 カテゴリとし，自動分類により振り分ける．また，新聞記事に書かれている事項が起こった場所，関連する場所と地図を紐付けすることで，地図上で新聞記事を選択する機能も実装する．

(2) キーワード検索に対するサジェスト機能

キーワード検索を行った際，検索力・語彙力の不足を補うために，サジェスト機能を実装することで，キーワードの入力補助を行う．

(3) 記事推薦機能

利用者が児童の場合，学年や興味などのユーザ情報をシステムに入力してもらい，それらの情報と他の小学生の興味情報，学習テーマなどを用いて，利用者である小学生の閲読意欲を掻き立てるような記事を推薦する．

利用者が教師の場合，社会科や総合学習などの授業で利用できる記事の推薦を目的としている．そのため，教科書の単元や学習指導要領から知識を抽出し，現在学習中の単元と関連する内容の記事を推薦する．

(4) 不適切記事のフィルタリング機能

自動収集した Web 新聞記事群に含まれる不適切記事をフィルタリングする．

以降，本稿では，不適切記事のフィルタリング機能と教師向けの記事推薦機能について概説する．

4. 不適切記事のフィルタリング機能

図 4 に本機能の概念図を示す．2.2 の調査により，不適切記事の約 50%は刑法で規定されている罪名を含むことがわかった．しかし，刑法に含まれない事故や自殺，いじめなどは漏れてしまう．そこで本研究では，2 段階のフィルタリングで対応する．具体的には，まず，刑法の罪名を利用したキーワードフィルタリングを行い，次に，SVM（Support Vector Machine）を利用し，キーワードフィルタリングで漏れた不適切記事をフィルタリングする．

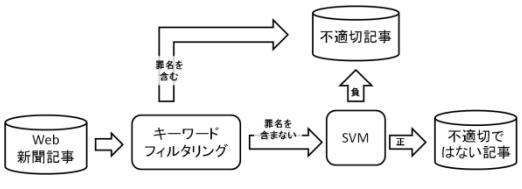

図 4　不適切記事フィルタリング機能の概念図

4.1 キーワードフィルタリング

1 段階目のフィルタリングであるキーワードフィルタリングでは，不適切記事とそれ以外の記事に分類する．判定に利用する刑法の罪名は，法なび法令検索[6]に掲載されていたものを利用する．

フィルタリングの手順を以下に示す．

(1) 自動収集した Web 新聞記事群から 1 記事を選出する．

(2) 罪名が記事中に含まれるか判定し，罪名が含まれていれば，不適切記事としてフィルタリングする．

フィルタリングされなかった記事が，次のフィルタリングの入力となる．

4.2 SVM によるフィルタリング

2 段階目のフィルタリングでは，キーワードフィルタリングで漏れた不適切記事をフィルタリングする必要がある．したがって，罪名以外の情報を素性に利用しなければならない．キーワードフィルタリングを通過した不適切記事を分析した所,火事や自殺など罪名を含まない記事であった．

そこで本研究では，不適切記事に頻出する単語に注目し，頻出単語と記事中に含まれる単語の共起の強さを素性として利用する．共起の強さを測る指標として，共起頻度，Simpson 係数．Jaccard 係数，コサイン距離，ダイス係数などがある．予備実験を行った結果，共起の強さ（共起度）を図る指標としてコサイン距離を用いて求める場合が最も精度が高かった．そこで，共起度を以下のコサイン距離で求める．

$$Cos(X,Y)=\frac{|X\cap Y|}{\sqrt{|X|\times|Y|}} \qquad (1)$$

ここで，X と Y は単語であり，$|X|$ と $|Y|$ は各単語のヒット数，$|X\cap Y|$ は X と Y の AND 検索のヒット数である．

以下に素性の選出手法を示す．

(1) 不適切記事を $d_{ni} \in D_N$, 不適切でない記事を $d_{pi} \in D_p$ とする．

(2) $d_{ni} \in D_N$ を形態素解析した後，品詞が名詞の単語 $w_j \in W_N$ が出現する文書頻度 $freq\_w_j$ をカウントする．

(3) $freq\_w_j$ の上位 n 件を不適切記事群 D_N における特徴語 $c_k \in C_N$ として抽出する．

(4) $d_{pi} \in D_p$ を形態素解析した後，品詞が名詞の単語 $w_i \in W_p$ を収集し，$c_k \in C_N$ と w_i のコサイン距離 $\mathrm{Cos}(c_k, w_i)$を求める．

(5) $\mathrm{Cos}(c_k, w_i)$の最大値を記事 d_{pi} における c_k の素性として利用する．

以上により，各 d_{pi} に対して n 次元の素性ベクトルが得られる．分類時にはキーワードフィルリングで非有害と判定された記事群を D_p として（4）と（5）を実行し，素性ベクトルを得る．

5. 記事推薦機能

本稿で提案する機能は，教師を対象とし，教科書の内容に関連する記事をランキング提示することで推薦する．さ

らに，推薦記事に対して適合性フィードバックを導入することにより，授業の内容に最適な記事を推薦する手法を提案する．なお，教科書は事前に電子化しておくことが必要となる．

5.1 推薦手法

推薦手順を以下に示す．

(1) 記事を推薦してもらいたい教科書の項を指定する．

(2) 該当ページに対して特徴ベクトル（ベース特徴ベクトル）を作成する．なお，単語は名詞のみ利用し，特徴量は TF-IDF を利用する．

$$TF\text{-}IDF = \frac{n_{ij}}{\sum_k n_{kj}} \times \log \frac{D}{d_{w_i}} \qquad (2)$$

ただし，特徴ベクトルを作成する際，項見出しは，重要な単語を含む可能性が高いので，見出しに含まれる単語の特徴量は 2 倍する．

(3) 新聞記事集合の各記事に対して，特徴ベクトル（記事特徴ベクトル）を作成する．なお，単語は名詞のみ利用し，特徴量は TF-IDF を利用する．

(4) 教科書の該当ページに出現する単語 w_i と新聞記事集合内の単語 w_j の共起度（Cos 距離）を求め，閾値を越える w_j をベース特徴ベクトルに追加する．

(5) ベース特徴ベクトルと各記事特徴ベクトルとの類似度 sim を求め，その値を基に記事をランキングして推薦する．

$$sim(\vec{X}, \vec{Y}) = \frac{\sum_{i=1}^{n} X_i \times Y_i}{\sqrt{\sum_{i=1}^{n} X_i^2} \times \sqrt{\sum_{i=1}^{n} Y_i^2}} \qquad (3)$$

(6) 目的の記事が推薦された場合，終了する．さもなければ，推薦記事の上位 n 件を 5 段階評価してもらう．

(7) ベース特徴ベクトルに，高評価であった記事に含まれる単語の特徴量を 2 倍，低評価であった記事に含まれる単語の特徴量を 0.5 倍する．

(8) (5) に戻る．

(6) 以降の適合性フィードバックにより，教科書の内容と関連する記事でかつ，ユーザの目的に適した記事が推薦可能となる．

6. 評価実験

4.で提案した不適切記事のフィルタリング手法，5.で提案した記事推薦手法の有効性を評価する．

6.1 不適切記事のフィルタリングの評価

新聞記事は，見出しとリード文が要約になっている[7]ことから，記事のタイトルと記事本文の 1 段落目に含まれる名詞の TF-IDF 値を素性として利用する手法（ベースライン）と比較することで，提案手法の有効性を評価する．

評価対象は，毎日.jp で 2012 年 1 月に掲載された 1,237 件の記事を利用する．学習データには，毎日.jp で 2011 年 9 月に掲載された記事 100 件を利用する．負例には犯罪やいじめなど小学生にとって不適切であると思われる記事（50 件）を，正例の学習データには，科学技術や地域情報など小学生にとって有益であると思われる記事（50 件）を使用する．なお，SVM には SVMlight（線形カーネル）を利用する．

表 1 に評価結果を示す．表 1 より，キーワードフィルタリングの場合，精度は高いが再現率が 0.45 となり，約半数の不適切記事をフィルタリングできない．SVM を用いた 2 段階フィルタリングでは，再現率が大幅に改善された．次に，ベースラインと提案手法を比較すると，提案手法の再現率が向上している．これは，素性として不適切記事で頻出する単語と共起度が高い単語を用いることで，不適切記事をフィルタリングできたと考えられる．しかし，精度はベースラインと比較して提案手法の方が低い．これは，ベースラインと比べて素性の次元数が非常に少ないため，分類器が非常に汎用的になってしまい広範囲の記事をフィルタリングしてしまったためだと考えられる．

表 1　フィルタリングの評価結果

	再現率	精度
キーワードフィルタリング	0.45	0.92
ベースライン（キーワードフィルタリング＋SVM）	0.61	0.51
提案手法（キーワードフィルタリング＋SVM）	0.89	0.28

6.2 記事推薦手法の評価

適合性フィードバックの有無による推薦結果の差を比較することで，提案手法の有効性を評価する．それぞれの結果に対して，3 名の被験者（大学学部生 2 名，大学院生 1 名）に 5 段階（5 : 非常に関連している，4 : ある程度関連している，3 : 関連している，2 : あまり関連していない，1 : 全く関連していない）で評価してもらう．なお，共起度の閾値は 0.4 として実験を行った．図 5 に評価結果を示す．図 5 より，評価値が 4 以上のものを見ると，適合性フィードバックを行わなかった場合は約 35%であるのに対して，適合性フィードバックを行った場合では，約 60%と評価値が高いものが大幅に増えたことが確認できる．このことから，適合性フィードバックを用いることにより，利用者の目的としている記事を推薦しやくすなることが判明した．

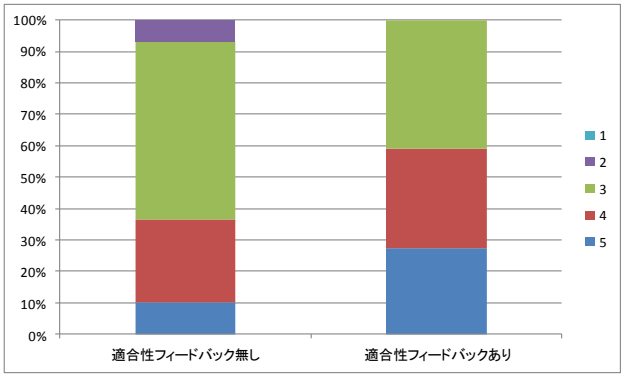

図 5　記事推薦手法の評価結果

7. おわりに

本稿では，不適切記事のフィルタリングに罪名を利用したキーワードフィルタリングと SVM を組み合わせた 2 段階の不適切記事のフィルタリング手法，教科書の内容でかつ，ユーザの目的に適した記事の推薦手法を提案した．

評価の結果，不適切記事のフィルタリング手法は，再現率は向上したが，精度は大幅に低下してしまった．記事推薦では，適合性フィードバックを用いることで，利用者が目的としている記事を推薦できることを示した．

今後は，フィルタリングの精度を向上させるために，素性を再検討する．また，記事の推薦では，教科書以外で個人の趣味・興味の情報から記事を推薦する手法について検討する．最終的に，記事選出支援システムの実現を目指す．

謝辞　本研究は，文部科学省科学研究費補助金（若手研究(B)22700813）の助成を受けて実施した

参考文献

1）教育に新聞を，http://www.nie.jp/
2）Fabrizio Sebastiani, “Machine learning in automated text categorization”, ACM Computing Surveys, Vol. 34, 2002.
3）　NIE に関する調査　第 5 回 NIE 効果測定調査　2010 年 7 月，http://nie.jp/research/pdf/re5_201007.pdf
4）福島、小原、須原、生田、コンピュータ＆エデュケーション,18,112－120,2005
5）よみうり博士のアイデアノート，http://www.yomiuri.co.jp/nie/note/kids/index.html
6）法なび法令検索，http://hourei.hounavi.jp/
7）記者ハンドブック第 9 版，共同通信社，2002.