

小学校における NIE 支援のための地図を用いたニュース閲覧システム  
- 県別の記事内容の分析とその活用法の検討 -

A News Browsing System using a Map for NIE Support in Elementary Schools  
- Analysis of Contents of Articles by Prefecture and its Utilization -

内山 豊<sup>†</sup> 黒田 晃史<sup>†</sup> 安藤 一秋<sup>‡</sup>  
Uchiyama Yutaka Akifumi Kuroda Kazuaki Ando

## 1. はじめに

近年、全国の小学校をはじめとした教育機関では、新聞記事を教材として活用する教育 NIE (Newspaper in Education) が実施されている。NIE を実施することにより、児童の読解力や社会への関心が高まるだけでなく、家庭内でも記事内容を話題にした会話も増加することで、コミュニケーション能力の育成などに効果がある[1]。

NIE では、各新聞社が発行する紙媒体の新聞を中心に利用するが、近年では、各社が Web 上で公開している Web ニュースも活用されるようになった。しかし、新聞記事や Web ニュース記事は、児童を対象として書かれていないため、児童が理解できない単語や表現などが存在する。したがって、新聞記事を読んでも内容を理解できないだけでなく、興味のある記事を探すことも困難である[2]。そこで、Web ニュース記事の閲覧をサポートするシステムがあれば、NIE の効果を高めることができると考える。

小学校での NIE では、地域に関連した記事を学習テーマとして取り上げ、身近な地域の特徴や出来事などを調査する実践例が多い[2]。また、小学校の中・高学年で、地図の読み方や都道府県の位置を覚えるなど、地図に関する学習を行う。

以上の背景の下、本研究では、小学校中・高学年を対象に、閲覧インタフェースとして地図を採用した Web ニュース閲覧システムの開発を進めている。本稿では、現在開発しているシステムの概要を説明し、現在検討している県別の特徴語を抽出・可視化することで、NIE や地域学習に活用する方法について述べる。そして、特徴語を抽出する手法について検討する。

## 2. Web ニュース閲覧システムの基本機能

本研究では、以下の機能を有する Web ニュース閲覧システムの開発を進めている。なお、現時点では、(1)から(3)の機能を実装済みである。

- (1) NIE に不要な記事のフィルタリング
- (2) カテゴリに基づく記事マーカーの色分け
- (3) 記事の絞り込み
- (4) 平易化した記事概要の表示

Web ニュース閲覧システムのメインインタフェースを図 1 に示す。インタフェースの画面中央には、Google Maps API[3]に基づく地図を表示し、ニュース記事から抽出した位置情報を基にマーカーを配置する。マーカーは、図 1 のように広域表示時には集約し、件数を表示する。また、図 2 のように、マーカーはカテゴリごとに色分けして配置する。地図の上側には、キーワード、期間、カ

テゴリに基づく絞り込み機能のフォームを配置する。また、地図上のマーカーがクリックされると、図 2 に示すように、「記事の見出し」「画像」「発行日」「新聞社名」からなる情報ウィンドウを表示し、同地点に紐づく記事リストを地図の下側に表示する。記事タイトルがクリックされると、記事概要等を表示する。

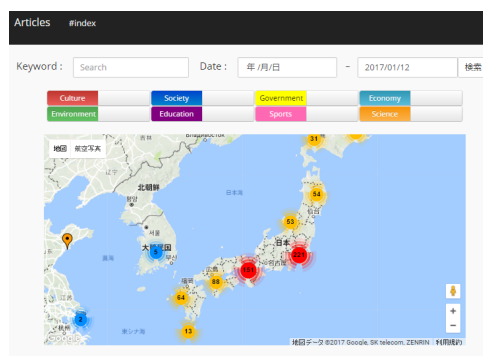


図 1 メインインタフェース



図 2 情報ウィンドウと記事見出しリスト

## 3. 県別の記事分析と活用法の検討

現在、ニュース記事を分析し、県別の特徴語を抽出・可視化することで、NIE や地域学習に活用する方法を検討している。ニュース記事は動的情報源であるため、ニュース記事から抽出した情報は、地図帳や教科書のような静的情報と異なり、鮮度の高い情報であるといえる。そこで、任意の期間における各県の特徴語を抽出・可視化することで、注目されている産業や特産品、文化などを把握する素材に活用できると考える。

これらの情報を提供することで、児童は、興味のある地域で、どのようなことが注目されているのかを把握でき、NIE での記事選択の手がかりに使えると考える。また、県別の特徴語は、記事推薦にも活用できると考える。さらに、地図帳などで使われている地域の特産品や産業などの静的情報と同時に可視化することで、地域理解を

<sup>†</sup>香川大学大学院工学研究科

Graduate School of Engineering, Kagawa University

<sup>‡</sup>香川大学創造工学部

Faculty of Engineering and Design, Kagawa University

深める学習支援が可能であると考えます。

以下、本稿では、県別の特徴語を抽出する手法について検討する。

#### 4. 特徴語の抽出

県別のニュース記事から特徴語を抽出する方法について述べる。情報源には、2017年12月中旬から5月中旬までの半年間で収集し43,537件の記事を利用する。

##### 4.2. 形態素解析

県別の特徴語を抽出するために、固有名詞を長単位で解析できるSudachiを利用する。本稿では、形態素解析結果のうち、特徴語を抽出するという観点から、名詞(数詞を除く)を特徴語の候補とする。

##### 4.3. 特徴語の抽出

特徴語の抽出には、以下の3手法を比較し、その性能を確認する。

###### ・TF-IDFに基づく特徴語抽出 (TF-IDF)

TFは、1都道府県の記事集合内の形態素 $w$ の大域的頻度とする。IDFは、その記事集合内において、 $w$ が出現する記事数をDFとして算出する。

###### ・改良IDFに基づく特徴語抽出 (TF-EIDF)

TFは、上記と同様で、IDFは、都道府県別の47個の記事集合のうち、 $w$ が出現する記事集合数をDFとして算出する。

###### ・TF-BDC[4]に基づく特徴語抽出 (TF-BDC)

Wangらが提案したTF-BDCは、エントロピーと形態素 $w$ が出現するカテゴリ内における出現頻度を組み合わせ、形態素 $w$ のエントロピーの平均を利用して重要度を算出する方法である。カテゴリ $C_i$ 内の記事に出現する形態素 $w$ の出現数を $f(w, c_i)$ とすると、 $bdc(w)$ は以下の式となる。なお、 $BH(w)$ は、すべてのカテゴリにおける形態素 $w$ のエントロピーの平均である。

$$bdc(w) = 1 - \frac{BH(w)}{\log(|C|)} \cdots (1)$$

$$= 1 + \frac{\sum_{i=1}^{|C|} \frac{p(w, c_i)}{\sum_{i=1}^{|C|} p(w, c_i)} \log \frac{p(w, c_i)}{\sum_{i=1}^{|C|} p(w, c_i)}}{\log(|C|)}$$

$$p(w, c_i) = \frac{f(w, c_i)}{f(c_i)}$$

特徴語抽出では、新聞記事に出現する一般的な形態素を除く。また、情報量の少ない一文字の名詞や地名と判断された名詞も除外する。地名は、複数の都道府県に出現するものもあるが、都道府県固有の地名が多く、特徴語になりやすいため除外する。

#### 4.4. 特徴語抽出実験

3手法の性能を確認するため、半年分の記事データから各県ごとの特徴語を抽出した。表1と表2には、香川県と福岡県の記事集合に対し、3手法で抽出された特徴語の上位10語を示す。

TF-BDC以外の手法では一般語が上位として抽出されていることがわかる。一方で、TF-BDCでは瀬戸大橋や福岡県の小倉城など、県の特徴語と言えるものが多く抽出されている。しかし、複合化すべき形態素やNIEに不要な形態素など、改善すべき点が見られる。そのため、今後はTF-BDCをベースライン手法として精度の向上を目指す。

表1 香川県における特徴語のTop10

|    | TF-IDF | TF-EIDF | TF-BDC |
|----|--------|---------|--------|
| 1  | 出身     | 警察      | 瀬戸大橋   |
| 2  | 警察     | 出身      | 香川大学   |
| 3  | 去年     | 瀬戸大橋    | 浜田     |
| 4  | 今年     | 今年      | 廃棄物    |
| 5  | 午後     | 去年      | JR四国   |
| 6  | 運転     | 午後      | 高松空港   |
| 7  | 瀬戸大橋   | これ      | 大翔     |
| 8  | 男性     | 選手      | 琴勇輝    |
| 9  | 確認     | 男性      | 水車     |
| 10 | これ     | 今月      | 養鶏場    |

表2 福岡県における特徴語のTop10

|    | TF-IDF | TF-EIDF | TF-BDC |
|----|--------|---------|--------|
| 1  | 警察     | 警察      | 豪雨     |
| 2  | 選手     | 去年      | 有明海    |
| 3  | 去年     | 今年      | 開門     |
| 4  | 女性     | 選手      | 九州電力   |
| 5  | 逮捕     | 女性      | JR九州   |
| 6  | 今年     | これ      | ソフトバンク |
| 7  | 午後     | 男性      | 工藤会    |
| 8  | 男性     | 午後      | 小倉城    |
| 9  | 影響     | 影響      | 土砂崩れ   |
| 10 | これ     | 逮捕      | 九州大学   |

#### 5. おわりに

本稿では、小学校でのNIE支援を目的としたニュース閲覧システムにおいて、これまでに収集した県別の記事集合から特徴語を抽出し、NIEや地域学習に活用する方法について述べた。そして、3手法を用いて特徴語を抽出する手法を検討した結果、TF-BDCに基づく手法の性能が高いことを確認した。

今後は、TF-BDCの改良や、他の指標に基づく特徴語抽出法を検討し、特徴語抽出の性能向上を目指す。そして、特徴語の可視化法についても検討する。また、提案システムのインタフェースと機能について、小学校の教師へのアンケート評価を行い、システムを改善する。さらに、当研究室で実施している平易化の研究や記事に対して補足情報を提示する研究などと組み合わせ、教育現場での実用を想定したシステムを構築し、総合評価を実施する。

#### 謝辞

本研究の一部はJSPS科研費16K00478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] 小林他, “小学生を対象とした新聞読解支援のための説明語抽出手法”, 情報処理学会研究報告, Vol.2013-CE-119, No.17, pp.1-6, 2013.
- [3] Google Maps API, <https://developers.google.com/maps/>
- [4] T.Wang, Y.Cai, H.f.Leung, Z.Cai and H.Min. "Entropy-based Term Weighting Schemes for Text Categorization in VSM," Proc. of the 27th Int'l Conf. on ICTAI, pp.325-322 (2015).