

高専関連報道記事を活用した活動情報の獲得と分析

小西 慶和† 藤田 直哉‡ 中島 陽子‡ 本間 宏利‡ 山本 和英*

† 釧路高専 電子情報システム工学専攻

‡ 釧路高専 情報工学科

* 長岡技術科学大学 電気系

yoko, honma@info.kushiro-ct.ac.jp

1 はじめに

2014年度現在，国立高等専門学校機構本部（以降，機構本部）のもと，日本全国に51の国立高専が存在する．各高専は，日々の教育や研究で得られた成果，地域貢献活動，学生の課外活動，学校行事などを地元の新聞，テレビなどのメディアや各種広報紙を利用して世の中に積極的に公表している．2004年頃からこれら各高専の公表記事を高専関連報道記事（以降，報道記事）として，機構本部が高専総合データベース KOALA (Kosen Access to Libraries and Archives) にて集約化サービスを開始した¹．このサービスは発信者である高専の担当事務職員が，報道記事を適宜要約して機構本部に投稿し，機構本部がそれらを取りまとめる形式で遂行されてきた．サービス開始当初は，しばらくは投稿数も少なく，有効的に活用されていたとはいえない状況であった．2009年頃から，各高専の啓発や意識の向上を促すことを目的に，これらの報道記事は定期的に（ほぼ毎週）メーリングリスト形式で全高専全教員に転送されるようになった．これ以降，登録データ数が右肩上がり増加していることが表1からも読み取れる．なお，メーリングリストにて転送される方式は2014年度から停止されて，現在は，各高専の更新情報のウェブページにリンクが張られた形式に変更された．

本研究では，2009-2013年までに配信された報道記事を対象として，それらを「学生活動」「地域貢献活動」「教育研究活動」「国際交流活動」「その他」の5つのカテゴリへの自動分類を行なう．また，分類後の報道記事を解析処理することで，全国の高専の研究や教育に関するトピックスの取得や，他校の学生活動の最新動向の把握，遠方の地域の高専の地域活動の実状，高専生のグローバルな貢献度など客観的に数値化

表 1: 年度別高専関連報道記事数

年度	報道記事数
2004	143
2005	473
2006	704
2007	900
2008	1,019
2009	2,537
2010	3,036
2011	3,498
2012	3,644
2013	4,283

してとらえることを可能とした．全国の高専の平均値や，特定の高専との数値を，本校のそれらと比較することで，今後の学校活動指針へフィードバックされるような，新たな課題が発見されることを期待して本研究を行なった．

2 研究手順

2.1 メール記事の自動分類

本研究では，最初に報道記事の「学生活動」「地域貢献活動」「教育研究活動」「国際交流活動」「その他」の5つのカテゴリへの自動分類を行った．学習器にランダムフォレスト [1] を使用した．分類に必要な教師データとして，各カテゴリごとに手動分類した50件ずつの記事を与えた．

自動分類結果の精度を2013年度の半年分のデータをサンプルとして導出した．その結果を表2に与える．

2.2 活動傾向の分析

つづいて，上記処理にて分類した各記事カテゴリごとに，それらに含まれる記事を利用して，活動傾向の

¹高専総合データベース KOALA, <http://www.kosen-k.go.jp>.

表 2: 自動分類結果の精度

記事カテゴリ	再現率	適合率	F 値
学生活動	0.792	0.645	0.711
地域貢献活動	0.600	0.522	0.558
教育研究活動	0.607	0.791	0.687
国際交流活動	0.607	0.436	0.507
その他	0.592	0.774	0.671

分析を行う。ある記事カテゴリ内の記事を各年度ごとに、形態素解析器 MeCab²を利用して、名詞、固有名詞、サ変名詞を抽出し、それらの語句を含んだ記事数の多いものから順に、その語句を列挙する。

年度に関係なく、多くの記事に出現する語句は、高専の特徴を表すものが並ぶと予想される。また、年度ごとにその語句が使われている記事数に大きな変動の見られる語句は、何らかの時代背景的、地域的な特徴を表すものであったり、精力的に取り組まれた活動であったと考えられる。よって、これらのような記事分析結果を活用して、全国高専の活動傾向を把握でき、今後の活動指針の参考にもなると期待する。

以降、紙面の関係上、分析対象を「教育研究活動カテゴリ」に限定して、論議をすすめる。

2.2.1 高専の基盤的研究

表 3 に、2011 年度と 2012 年度の教育研究活動カテゴリの報道記事を対象に、それが含まれる記事数の多い順に語句を提示する。

表 3: 2011-2012 年度の高使用頻度語句

2011 年		2012 年	
本校	47	本校	47
研究	45	研究	42
高専	42	高専	36
学生	32	学生	33
技術	23	講演	25
工業	22	技術	17
講演	21	専攻	17
高専機構	19	実用	16
独立行政法人	18	工業	15
対策研究	17	科学技術振興	15
実践教育	16	専門	12
産学	16	産学	12

表 3 に現れている語句は、現在の高専教育研究の基幹・基盤活動に深く関係する語句と考えられる。しかし、表 3 は、高専の教育研究活動の大枠を表しているものの、この結果だけでは、研究活動傾向や、地域ごとの特色、萌芽的・独創的な流行が見えにくい。

²<http://mecab.sourceforge.net/>.

2.2.2 高専の短・中期的な研究活動

表 4 は、年度ごとで、それらを含む報道記事の投稿数に、大きな差異が見られた語句の一覧である。

表 4: 使用記事数の差が顕著な語句

ワード	2009	2010	2011	2012	2013
対策研究	0	1	17	0	0
被災	2	1	11	2	3
独立行政法人	3	2	18	5	2
国立高専機構	0	4	19	5	5
大震災	0	1	32	23	7
エネルギー	6	10	30	13	8
防災	11	6	29	17	14
連携	58	103	127	65	61
実用	6	11	7	16	10
経済産業	1	10	2	0	3
講演	13	16	21	25	20
津波	3	8	27	12	4

表 4 は、ここ 5 年間で、各高専が精力的に研究が行ってきた研究分野やその背景を表している。「被災」、「大震災」、「津波」などのワードは、2011 年の東日本大震災による背景を示している。また、その直後から「対策研究」、「エネルギー」、「防災」などの語句が増加していることは、震災対策、エネルギー対策に関する研究が、各高専でなされてきていることを意味している。また、「独立行政法人」、「国立高専機構」、「連携」、「経済産業」などの語句も、ここ数年の高専の研究活性化を推進する活動が活発になってきている背景を顕著に表している。

表 4 形式のデータを分析することで、世の中の要望、期待に即座に対応する重要性が再認識でき、今後の短・中期研究計画の参考となる。

2.2.3 高専の萌芽的・挑戦的な研究活動

表 5 は、記事数は多くないものの、年度ごとに、着実にそれらを含む報道記事数が増加している語句の一覧である。

この結果では、「若手」、「子ども」、「リケジョ(女性)」の語句を含む、人的ターゲットを限定した報道記事と、「人材育成プログラム」、「イノベーション」、「グローバル」、「養殖」などの革新的な語句を含む記事の投稿が着実に増加していることがわかる。投稿数はそれほど多くないものの、これらをテーマとした研究増加の風潮は、我々自身も肌で感じている。

よって、表 5 形式のデータを分析することで、今後の大きな社会の要求、技術者教育の変遷、最先端研究

表 6: 地域別の記事投稿数

語句	北海道	東北	関東	東海	近畿	中国	四国	九州
エンジニア	9	0	1	0	6	0	2	0
人材育成プログラム	3	0	2	1	0	0	0	0
若手研究	3	1	2	0	0	0	0	0
子ども	2	0	0	1	0	0	1	0
養殖	8	0	0	8	0	0	0	0
イノベーション	12	0	3	2	4	1	0	1
グローバル	2	0	0	1	0	0	1	0
カキ	9	0	0	9	0	0	0	0
リケジョ	0	0	1	0	6	0	2	1
計	50	1	8	23	10	1	4	2
		2%	16%	46%	20%	2%	8%	4%

表 5: 使用記事数が増加している語句

語句	2009	2010	2011	2012	2013
エンジニア	4	6	6	7	9
人材育成プログラム	0	0	1	1	3
若手研究	1	1	1	2	3
子ども	0	1	2	2	2
養殖	0	1	6	8	8
イノベーション	3	7	12	12	12
グローバル	0	2	2	2	4
カキ	0	0	4	6	9
リケジョ	0	0	0	1	2

の流れを取り入れた, 中・長期的な研究計画指針を決定する材料になりえると思われる。

2.2.4 教育研究活動の地域特色

表 5 の示した記事数は多くないものの, 年度ごとに使用記事数が増加している語句の一覧であった。ここでは, その中の 2013 年度の投稿記事を対象として, それらの報道記事を投稿した高専を地域別に分類して集計した。その結果を表 6 に示す。

表 6 より, これらの革新的・挑戦的な研究は, 関東地区の高専が全体研究の約半数の先駆者的な役割を担っていることがわかる。これに東海地区の高専が全体の 20% の研究に携わっている。反対に, これらの地域から遠方の地域に位置する高専になるほど, これらの報道記事数, すなわち, 活動は極端に少なくなっている。残念なことに, 釧路高専の北海道地区はほとんど貢献していない。これらは, 大きな工業都市に隣接し, 大学, 大企業との関係が近い高専ほど, 世間の要求や最先端技術への挑戦に携わりやすい環境にあるということを示しているかもしれない。

3 おわりに

本研究では, メール転送される高専関連報道記事を利用した, 全国の高専の活動情報の獲得と分析を行った。入手した報道記事は, ランダムフォレスト学習器を使って, 5 つの記事カテゴリに自動分類させた。また, 分類後のカテゴリ別の記事データを利用して, 年度時系列での活動傾向分析を行った。これにより, 全国高専の「学生活動」「地域貢献活動」「教育研究活動」「国際交流活動」のそれぞれにおける活動傾向を把握できることの可能性を示した。

本紙では, 紙面の関係上, 教育研究活動に限定して活動傾向分析結果の一例を示したが, 発表時には別紙にて, より詳細な他カテゴリ記事の活動傾向分析結果の公表を予定している。

謝辞

本研究は高専-長岡技科大共同研究の助成 (2014) を受けて行なわれました。

参考文献

- [1] L. Breiman. Random forests. *Machine Learning*, Vol. 45, No. 1, pp. 5–32, 2001.