「第21回年次大会予稿」

学術研究のメディア報道における定量的調査研究

ープレスリリースと新聞報道の関係ー

A Quantitative Investigation of Academic Research Findings in

Media Report Articles

- Relationship between press release and newspaper article -

西澤正己*, 孫媛

Masaki NISHIZAWA*, Yuan SUN

国立情報学研究所

National Institute of Informatics

〒101-8430 東京都千代田区一ツ橋2-1-2

E-mail: nisizawa@nii.ac.jp, yuan@nii.ac.jp

大学等の研究成果はどのように一般に報道されているのかを、研究機関別や分野別に調査する。今回は、プレスリリースされた発表と全国紙の新聞記事の関係を中心に、採用率や新聞社の対応の違いなど、調査した結果を報告する。

It is investigated how results of academic research in universities are reported towards the public. This time, the difference in correspondence of an adoption rate and a newspaper publishing company centering on the relation of the newspaper article of the announcement by which Press Releases were carried out, and a national paper, etc. report the investigated result.

キーワード: 大学, 研究成果, 報道, プレスリリース, 新聞

Keyword1, University, Research output, Press report, Press release, Newspaper

1 はじめに

我々はこれまでに学術研究の動向や分野間の関係、産学連携の状況を調査してきた[1][2]。これらの研究の中では研究成果や研究動向の情報として学術論文や研究資金(科学研究費補助金等)のデータベースを利用してきた。しかし、社会一般に研究動向が広報されるのは新聞やニュース、インターネット等のメディアである。本研究では新聞記事、ニュース等の検索サービスを用い、一般報道とこれまで調査してきた学術動向との関係を調査研究する。

第一段階としては、特に「大学」に注目して、どの程度新聞やニュースリリースに大学関係の報道があるのかについて予備調査した[3]、本研究ではこれをさらに発展させ、プレスリリースと新聞報道の関係について報告する。

2 データ

2.1 プレスリリース

プレスリリースは発表する機関側から新聞メディア等への能動的な発表形態であり、「大学」関連のプレスリリースにおいては、大学が成果等を発表したもの、企業との産学連携や提携、その他、企業の製品が大学などに納入されたというような記事も含まれている。

日経テレコン21[4]での検索では、見出し記事の他、記事本文の検索が行われる。また、検索は本文や見出しに対して付けられたキーワードに対して行われているので、「大学」で検索した場合は、本文には「東大」しかなくとも「東京大学」がキーワードに含まれており、機関名に関しては、省略形でもほぼ検索にはヒットするようである。

今回の分析では、2005年から2011年の間に発表されたプレスリリースに対して「大学」を検索語にしてヒットした8,648件の見出し、および本文の記事を調査した。

2.2 新聞記事

新聞記事については、日経テレコン21の検索を併用しつつ、日外アソシエーツから発売されている新聞記事データ集[5]を使用した。ここでは、朝日、読売、毎日の新聞記事データが利用できるが、今回の分析では朝日新聞記事データ集(2007年から2011年版)、および読売新聞記事データ集(本社版、地方版、2007年から2011年版)を使用した。

読売新聞に関しては、このデータ集には2011年版で313,205件の記事が収録されている。日経テレコン21の検索結果では313,210件となっており、ほぼ同じ件数が収録されている。しかし、朝日新聞に関しては、2011年版においては、140,170件の記事しか収録されておらず、日経テレコン21の結果の約半数である。このデータはコーパス目的であるため、統計分析にはやや注意が必要かもしれない。

3 分析手法および結果

3.1 プレスリリースの分析

2.1で述べたように、「大学」の検索語でヒットした2005年から2011年に掲載されたプレスリリースの見出し8,648件について分析をおこなった。

日経テレコン21のプレスリリースの見出しでは、表1に示すように、はじめに発信主体の機関名があり、次に内容説明がある。これは、日経プレスリリースが、下記機関からのプレスリリースの見出しを編集して、独自

表1：プレスリリース分析における機関抽出と見出し中の最終キーワード例

記事番号	関連大学1	機関数	最終キーワード	本文見出し
1	名古屋大学	1	成功	JSTと名大、試料1個で熱電材料の性能を最適化することに成功
3	京都大学	1	発見	京大など、ホップの苦味成分を作る遺伝子を発見
6	慶應義塾大学	2	応用	資生堂、慶大や九州大と共同で「D-アミノ酸」の分析技術を神経難病の解明に応用
8	東京大学	1	公開	凸版印刷と東大、次世代型的美術鑑賞システム「デジタル展示ケース」を東京国立博物館で公開
9	理化学研究所	2	解明	理化学研究所と東大、嗅覚神経細胞分化の新しいメカニズムを解明
15	千葉大学	2	解明	千葉大と東大など、ホランド型型酸化物が強磁性を保ったまま金属から絶縁体になるしくみを解明
16	北海道大学	1	解明	北大、オートファジーの活性化にかかわるタンパク質の構造を解明
86	東北大学	1	合意	東京エレクトロン、東北大学とスピントロニクスメモリーの集積化技術などで共同開発に合意
95	長岡技術科学大学	1	構築	NEC、長岡技術科学大学の統合図書館システムをプライベートクラウドで構築
206	東京大学	1	締結	東大と水産総合研究センター、連携協力推進で協定を締結

に見出しを付けているものである。この見出しの最終キーワード（動詞）を用いることによって、発表内容が研究成果なのか協定の締結等であるかの分類をすることができる。また、見出しではほとんどの大学名は省略名で書かれている。本文に関しては内容を印刷形態で利用でき、内容の詳しい確認や、今後の分析に利用する予定である。

見出しの分析に関しては、検索結果を入力した上で、これまで行ってきたキーワード分析[6]と同様に、形態素解析システム茶筌を用いてできるだけ短い複合語になるようにキーワード抽出をおこなった。これらのキーワードを用い、我々がこれまでに開発してきた機関名の名寄せテーブルと照合することにより機関名に相当するキーワードが含ま

れているプレスリリースの見出しを抽出した。また、同時に見出しの最後のキーワードも抽出している。表1はこれらの処理を行った結果、名寄せされた機関名の最初のもの、機関数、最後のキーワード、元の見出し文の例を示したものである。最終キーワードの詳しい出現頻度等は前回の報告[3]で紹介している。

これらの処理をした結果、「大学」を検索語にしてヒットした8,648件中、見出し中にも名寄せデータベースに含まれる機関があった記事数は3,811件であった。また、機関数は延べ4,366機関である。これは、名寄せテーブルが論文の所属機関の名寄せ用であるため、大学名等はほぼヒットするが、ベンチャー企業等、論文をほとんど出さない機関

表2：大学関連のプレスリリースにおける最終キーワードの見出しに現れる機関セクタ別出現頻度

ekw分類名	SECT	2005	2006	2007	2008	2009	2010	2011
研究・開発	Univ	67	80	142	185	184	299	419
	Corp	165	150	170	181	179	180	195
開始	Univ	61	54	62	66	58	51	43
	Corp	210	211	201	188	212	185	172
連携	Univ	40	35	48	41	30	24	32
	Corp	118	94	123	108	94	76	80
提供・導入・構築	Univ	31	28	25	29	36	38	55
	Corp	301	281	293	292	322	297	282
発表	Univ	6	9	11	16	21	22	34
	Corp	71	103	93	91	86	102	114
その他	Univ	6	4	6	4	8	6	15
	Corp	102	81	88	80	90	66	89

はデータベースに含まれないためである。しかし、見出しに大学名が含まれる大学等が主体となって発表したプレスリリースについては、この方法でも十分抽出できるのではないと思われる。

機関の名寄せテーブルでは「大学」、「企業」、「公的研究機関」等、セクター分類もできるようになっており、後の分析にも利用している。表2にはプレスリリースの見出し記事の最終キーワードを、「研究・開発」に関するもの、事業等の「開始」に関するもの、機関間の「連携」に関するもの、製品等の「提供・導入・構築」に関するもの、社会調査結果等の「発表」に関するもの、およびその他の6つに分類した。この分類別およびプレスリリースの見出しに現れた機関のセクター別の記事件数を示す。プレスリリースの記事は、

「大学」で検索してヒットした記事なので注意は必要であるが、企業セクター(Corp)と大学等研究機関セクター(Univ)の特徴が現れており、特に近年、Univセクターの「研究・開発」に関する記事が増えている事がわかる。

プレスリリースと新聞記事の内容一致を調べるため、プレスリリースの見出し記事に含まれる機関名および、発行日付を重視して分析する。

3.2 新聞記事の分析

本報告では、朝日新聞と読売新聞の記事データを用いて分析する。朝日新聞記事データに関しては、記事が掲載された版や面、分類記事の情報も含まれているが、キーワードは付与されていない。今回は、まず記事見出しを用いて分析をおこなう。プレスリリースの

表3：新聞記事数と記事中に出現した機関名のセクター別記事数

表3.1：朝日新聞見出し

year	朝日記事数	機関特定記事	Univ	Corp	Gov	Pub
2007	153540	35879	1868	16689	16654	443
2008	148317	34257	1747	15368	16450	518
2009	143553	34068	1739	14856	16996	412
2010	141411	34023	1596	13342	18727	387
2011	140170	33289	1557	13391	17677	506

表3.2：朝日新聞本文

year	朝日記事数	機関特定記事	Univ	Corp	Gov	Pub
2007	153540	96196	14855	79873	26703	4500
2008	148317	93875	14529	77611	26400	4789
2009	143553	90603	13913	75297	25086	4357
2010	141411	87300	13065	71880	25014	4213
2011	140170	86088	12816	71632	22819	4504

表3.3：読売新聞付与キーワード

year	読売記事数	機関特定記事	Univ	Corp	Gov	Pub
2007	343140	226009	26500	180236	82424	9482
2008	333193	217637	25339	172357	78287	9678
2009	316833	155887	23231	128382	67217	8279
2010	310750	195106	23365	150758	71470	8358
2011	313205	198811	23312	155206	70828	7678

分析と同様に見出し記事について、形態素分析システムを用い、キーワードに分解し、名寄せデータベースと照合したうえで、含まれる機関名を記事ごとに抽出した。

読売新聞については、本文のキーワードが付与されている。よって、機関名の抽出にはこのキーワードを用い、名寄せテーブルと照合して記事ごとに含まれる機関名を抽出した。表3は、朝日新聞の見出し（表3.1）および本文（表3.2）、読売新聞の付与されたキーワード（表3.3）を用いて、記事中出现する機関のセクター別の記事件数を示したものである。大学セクターが現れるのは、朝日新聞の見出し中で全体の1.2%程度、本文で9.5%程度、読売新聞のキーワード中でも9.5%程度である。

全国紙等の新聞記事はそのほとんどが社会、政治、経済、国際、スポーツ等のニュース記事であり、大学の研究成果が報告されるのはごくわずかである。

3.3 プレスリリースと新聞記事の相関

プレスリリースと新聞記事の内容の一致を調べるために、プレスリリースの記事見出し、新聞の見出しのそれぞれを前章の方法で、キーワードに分解し、コサイン距離[6]を求めた。ここで、キーワードの頻度の重みは

表4：コサイン距離と一致不一致頻度

Cos. Dist.	一致	不一致	比率
0.1	6	2290	0.26%
0.2	62	3229	1.88%
0.3	68	475	12.52%
0.4	76	114	40.00%
0.5	41	17	70.69%
0.6	29	1	96.67%
0.7	9	0	100.00%
0.8	4	0	100.00%
0.9	1	0	100.00%

$\log(N)$ としている。それぞれのキーワードの一致の頻度からコサイン距離は求めるが、組み合わせの数が多くなりすぎて、計算時間がかかるため、まず、それぞれの記事中出现する機関名が同じものに対して、また、プレスリリースが出されてから新聞記事が出るまでの日時差が-30から+60日までの間の記事について計算をおこなった。

図1に朝日新聞の見出しとプレスリリースの見出し記事間のコサイン距離の頻度分布を示す。内容の一致、不一致は、コサイン距離の値や記事間の日時差を参考に目視で内容の確認を行った。ただし、コサイン距離が0.2以下のものについては、日時差が-5から+10までの物しか確認をしていない。また、機関名が一致している物だけを計算しているので、コサイン距離が0.2以下になる物は少なくなる。この結果、プレスリリースの見

図1：コサイン距離の頻度分布（左：内容が一致した記事、右：不一致）朝日新聞見出し。

図2：プレスリリースと新聞記事間の日時差と一致、不一致頻度の関係（右：5日、右：1日刻み）

出しと朝日新聞の見出し間で一致すると確認された記事は264件であった。

表4には、コサイン距離が0.1ごとの、一致した頻度と不一致の頻度、その比を示している。コサイン距離が0.6以上ではほぼ100%が一致しているが、数は少なく、一致の頻度のピークは0.4付近にあり、このときの一致率は40%以下であるので、コサイン距離のみでは仕分けは難しい。

図2には一致、あるいは不一致のプレスリリースと新聞記事の記事間の日時差と頻度を示している。日時差が小さい（0から+2）までの間で、一致度が急峻なピークを迎えている事がわかる。このことより、コサイン距離と日時差を最適化すれば、かなりの確率で

内容が一致する記事を見つける事ができるであろう。また、今回は見出し記事のみの結果の報告であるが、本文を用いた計算もほぼ終わっているので、新聞記事の本文とプレスリリースの見出し記事を使うとさらに良い結果が得られるのではないと思われる。

3.4 一致記事の様相

表5には、朝日新聞見出しとプレスリリース見出し間の一致記事264件（表5.1）について、プレスリリースの最終キーワード（動詞）6分類別、および発表機関のセクター別頻度別、年別に示した。また、読売新聞見出しとプレスリリース見出し間の一致記事506件（表5.2）について、同様に示している。一

表5：新聞記事とプレスリリースの一致記事のセクター、記事分類別の記事数

表5.1：朝日新聞見出し-PRL見出し(264件) 表5.2：読売新聞見出し-PRL見出し(

ekw分類名	SECT	2007	2008	2009	2010	2011
研究・開発	Univ	19	16	9	23	27
	Corp	7	1	2	4	4
開始	Univ	5	6	1	1	4
	Corp	3	11	4	5	7
連携	Univ	4	5	4	2	3
	Corp	5	14	5	3	2
提供・導入・構築	Univ	0	1	3	1	2
	Corp	5	4	11	5	8
発表	Univ	1	1	0	2	3
	Corp	4	2	1	0	4
その他	Univ	0	0	0	1	2
	Corp	7	2	1	2	1

ekw分類名	SECT	2007	2008	2009	2010	2011
研究・開発	Univ	26	33	20	49	53
	Corp	16	9	11	15	19
開始	Univ	7	7	2	6	7
	Corp	15	13	7	19	18
連携	Univ	10	8	5	3	11
	Corp	17	19	4	4	11
提供・導入・構築	Univ	0	4	3	1	4
	Corp	7	11	7	4	14
発表	Univ	3	2	2	3	4
	Corp	4	4	2	2	8
その他	Univ	0	0	2	1	9
	Corp	8	5	5	5	14

致件数は朝日新聞と読売新聞で大きく2倍ほど差がある。これは、朝日新聞の記事数の少なさ、および、機関名の抽出方法の違いが大きく影響している。朝日新聞に関しても、機関名を見出しからではなく、本文から抽出した場合は、抽出機関数が3倍近くに増える(表3)ことから、本文との一致を見るとこの差はかなり縮まるのではないかと思われる。これは、間もなく結果が出るので、確認できるはずである。

どちらの結果からも、Univセクターの「研究・開発」が近年伸びているように見える。前回の結果から、近年のUnivセクターのプレスリリースの伸びが見て取れた。これに対応して新聞記事も伸びているように見える。これは、大学機関の積極的な広報が新聞メディアを通して広く一般に広報できていることを示すものであろう。さらなる分析がひとつである。

4 まとめ

大学等の研究成果はどのように一般に報道されているのかを、研究機関別や分野別に調査することを目的に、プレスリリースや全国紙の新聞記事を中心に予備調査した。今回は、朝日新聞の見出し記事とプレスリリースの見出し間で264件、読売新聞の見出し(機関抽出は付与キーワード)とプレスリリースの見出し記事間で506件の一致記事が見つかった。プレスリリースの発表機関セクターと記事種分類による分析では大学関係機関の「研究・開発」関連の記事の一致、すなわち、プレスリリースの新聞記事への採択が増えている結果となった。新聞社のデーによる違

いが大きいと、さらなる確認がひとつであり、新聞記事本文を用いた結果がひとつである。また、プレスリリースと新聞記事の一致を発見するための技術、コサイン距離と掲載日時の差の最適化等、技術の向上も今後の課題となる。

プレスリリース本文には研究結果が発表された論文情報が掲載されており、この情報の抽出も今後の課題である。今後はこの調査を基礎にして、他の新聞記事等も調査し、科研費や学術雑誌の論文数等との比較を行っていく予定である。

参考文献

- [1] 西澤 正己、孫 媛, 「キーワード分析による環境関連研究の動向調査」, 情報知識学会誌, Vol. 20, No. 2, pp. 155-162, (2010).
- [2] 柿沼澄男, 孫媛, 西澤正己, 大山敬三, 根岸正光, 「産学連携に関する Web 情報の分析 -大学サイトの事例研究-」, 情報知識学会誌, Vol. 19, No. 2, pp174-178, (2009).
- [3] 西澤 正己、孫 媛, 「学術研究のメディア報道における定量的調査研究」, 情報知識学会誌, Vol. 22, No. 2, pp. 138-143, (2012).
- [4] 日経テレコン21: <http://t21.nikkei.co.jp>
- [5] 日外アソシエーツ, 学術研究・開発研究のための言語資源コーパス:
<http://www.nichigai.co.jp/sales/corpus.html>
- [6] 藤垣裕子、平川秀幸、富澤宏之、調麻佐志、林隆之、牧野淳一郎 著, 「科学計量学入門」, 丸善, (2004).