

# 自然言語で書かれた複数の情報リソースの 時系列解析による市場探索手法の提案

上田 紗綾 宮前 義範 奥 良彰 中原 健

ローム株式会社 研究開発センター 融合技術研究開発部

{saya.ueda, yoshinori.miyamae, yoshiaki.oku, ken.nakahara}@dsn.rohm.co.jp

## 概要

本手法は、研究開発のテーマについてその市場性を分析する手法を提案する。研究開発テーマの代表的な単語を対象として、技術ニュースや特許データベースなど性質の異なる複数の自然言語で書かれた情報リソースを入力とし、Word2vecによるホットワード解析を実施する[1]。市場規模などまだ統計的な数値データがはっきりとしない分野や、その市場において独占的な分野があり他の分野が見えにくい状況においても、客観的かつ短時間に市場を俯瞰することを目指す。

解析テーマとして、酸素および湿度センサに関する市場分析を行った。どちらのセンサも、その既存市場は自動車産業が大きなシェアを占め[2][3]、それ以外のニッチな市場は見えにくくなっている。本手法により、技術ニュースサイトからは可能性のある市場のホットワードを、特許データベースからは、その市場と関連性の高い技術を取り出すことができた。

## 1 背景

研究開発テーマおよび事業成功確率の向上は喫緊の課題である。例えば、経済産業省が取りまとめた製造業別研究開発の成功確率のうち、「基礎研究が事業化に結びつくもの」は平均で 12.3%と、低い値にとどまっている[4]。このため、どの研究機関が有望であるかについて、国は研究開発に係る無形資産価値の可視化についての議論を進めている状況である[5]。しかしながら、研究開発領域、「知」の価値を評価する際の情報の妥当性について、該当する領域範囲での経験がなければ研究開発テーマや事業を適切に評価することができないことが課題となっており、またその評価手法については定式化することが望まれている。

一方で市場がまだ形成されていないことが多い研究開発のテーマについては、定量的なデータが容易

に手に入らないため統計的な手法を使うことができずに、各種定性的な情報を使用する必要がある。定性的な情報を解析するためのソフトウェアやコンサルティングが増えてきており研究も進んでいるが、特許や論文といった単一の情報源を処理する例がほとんどであり、例えば、ニュースなども含めた上での、複数の情報を組み合わせた評価はほとんどない。これは複数の媒体から得たニュースの評価を行おうとすると、質の異なるニュース自体を評価する必要があり、尚且つ、その重要度の判断を人が行なわなければならないためである。限られた時間の中でこれらの判断を適切に行うことは難しいのが現状である。

## 2 関連研究

売り上げ規模など市場における客観的な情報を使わない手法としては、大量のテキストデータの解析手法として、新規性の評価[6][7][8][9]、概要の要約[10]がある。また、特許をターゲットとしたものでは、記載技術の価値判断および時系列解析を用いた価値判断手法として、特許の引用数を使用して対象とする技術の価値を判断する手法[11] や、SCDV (Sparse Composite Document Vectors)による文書ベクトルと財務諸表のデータを使用して評価する手法[12]、論文本数を用いて評価する手法[13]がある。しかしこれらの手法は、あくまでも単一の情報リソースから分析したものであって、複数の情報リソースからデータを得て多角的に評価したものではない。

## 3 提案手法

本研究は、複数の媒体からデータ収集を行い、かつ技術者 1 名に対して事前ヒアリングを実施した。技術者より、酸素センサと湿度センサの想定している市場は「1.車載(自動車産業)、2.医療、3.産業機器、4.家電、5.モバイル、6.IoT(Internet of Things)」との回答を得た。また、3.1-3.3 でデータセットの収集方法及び解析手法を述べる。

### 3.1 解析対象とする媒体

本研究で対象としたニュースは「日経 TechFind[14]」「Anews[15]」「Share research[16]」の3つの媒体ごとに得たものである。異なる3つの媒体の検索結果を自然言語処理によって評価し、比較する。まず、日経 TechFind は技術的文章に強みがあり、独自に、これから伸びる技術領域についてユーザーが簡単に理解できるような工夫がされている。次に Anews は技術サイトのニュースの中から、研究開発部門に所属するメンバーの好みや反映されたサイトを選択できるため、その部門の嗜好を反映させることができる。最後に Shareresearch は特許に特化されている媒体であり、ほかの2つの媒体と比べて学術色が強い。

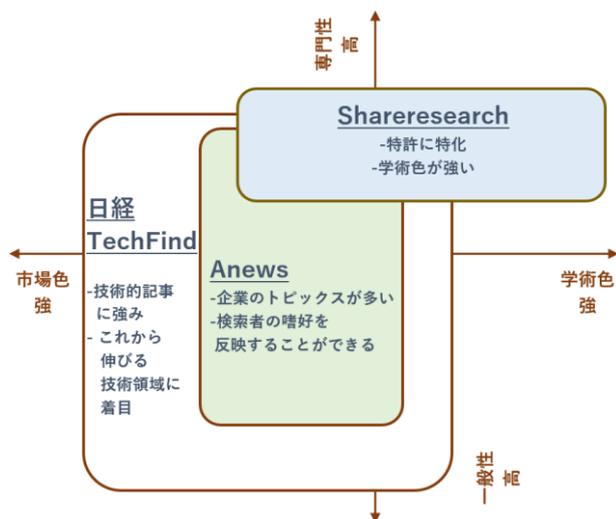


図1 対象とする3つの媒体について

### 3.2 解析対象とするデータ

データ範囲は、「日経 TechFind」は2017年から2022年、「Anews」は2019年から2022年(のぞく2020年)、「Shareresearch」2017年から2021年のものである。

### 3.3 解析フロー

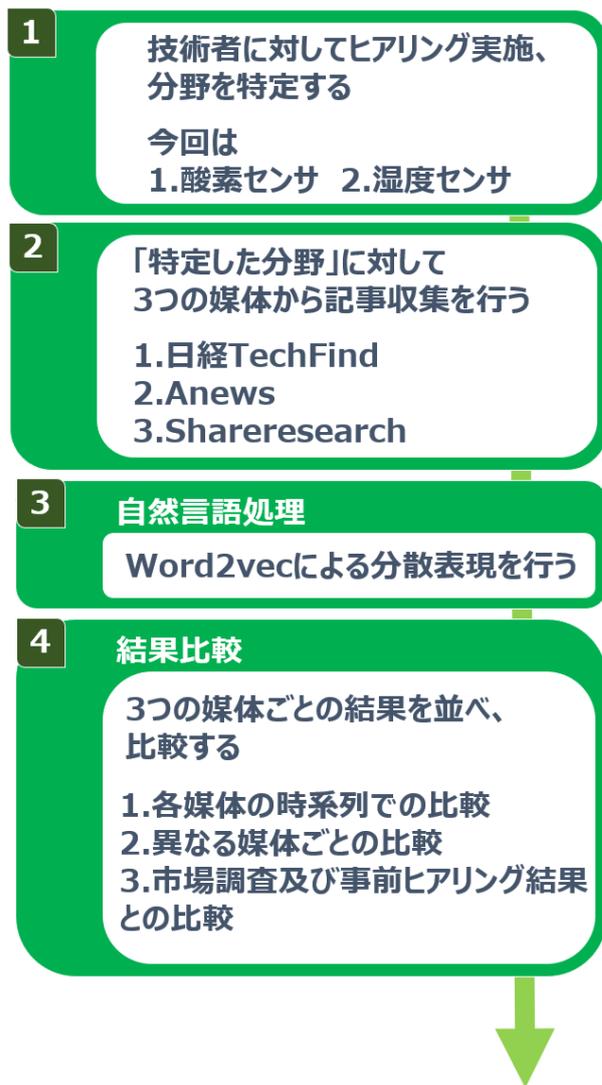


図2 本手法のフローチャート

図2に本研究のフローチャートを示す。技術者に対してヒアリングを実施し、どの分野について自然言語処理をしたいのか特定をする。次に、特定した分野に対して3つの媒体から記事収集を行う。そして、Word2vecにおける「skip-gram」[17]によりベクトル空間を作成し、t-SNE[18]による次元圧縮後、DBSCAN[19]によるクラスタリングを行った。特許だけに使用した例としては[1]があるが、今回は3つの媒体に対しても同じ手法を適用する。最後に、結果の比較を行う。

## 4 結果

### 4.1 「酸素センサ」の解析結果

酸素センサの解析結果を図3に示す、まず「日経TechFind」は、2019年度は「血中」「溶存」「無菌」、2020年度は「血中」「ウェルネス」「計測」「症状」、2021年度は「リハビリテーション」「感度」、2022年度は「代謝」「精度」という単語の出現が確認できた。これらの単語からは医療分野に関して酸素センサの適性を示すものである。一方「Anews」では「光線」「破断」といった各研究開発センターのメンバーが興味を持っている分野の単語が集まった。

「Shareresearch」に関しては「冷蔵庫」「NO<sub>x</sub>(窒素酸化物)」といった酸素センサが活用されるであろう技術的な領域に関する単語が導出された。(ここで述べた単語については図3で赤字表記を行った。)

これらの結果から、車載やIoTといった研究開発センターのメンバーの志向以外に、医療分野や環境系への市場への発展も期待できるという結果を得ることができた。

### 4.2 「湿度センサ」の解析結果

湿度センサの解析結果を図4に示す。「日経TechFind」は、2017年度は「環境」「照度」「気圧」、2018年度は「温度」「照度」「濃度」、2019年度は「温度」「照度」「濃度」、2020年度は「土壌」「日照」という単語の出現が確認できた。これらの単語からは農業分野に関して湿度センサの適性があるのではという事を示すものである。一方「Anews」は、「高機能」「コンセプト」「組み合わせ」「全体的」といった、システム全体に対してメンバーが興味を持っているということが判明した。「Shareresearch」に関しては、「アウトドア」「スプリンクラー」など、湿度センサの使用用途として複数の関連単語を得た。(ここで述べた単語については図4で赤字表記を行った。)

これらの結果から、酸素センサと同様に、研究開発センターのメンバーの志向以外に環境やシステム、特定の市場に関する知見を得ることができた。

	2017	2018	2019	2020	2021	2022	まとめ
日経Tech Find	転用 自動 起因 低温 インテグ レーション	濃度 燃焼室 センサー	血中 容器 検出 溶存 予防 無菌	推定 飽和 血中 ウェルネス 自覚 計測 症状 重症	予定 リハビリテ ーション 地球 年間 農業 感度	水素 導入 代謝 精度 酸性	医療分野
	項目なし	項目なし	リモートセン 光線 スケラビ 汎用的 可逆	転移 鏡像 バルク 曲面 可溶	水面 強度 等方 依存 破断	観察 導出 受容体 細胞核	メンバーが 関心ある 分野
	pressur e pemea- tion re- frigerat or	waste	gases ab- sorber	NO <sub>x</sub> normo- xia	air cannis- ter	項目なし	技術領域

図3 ホットワード「酸素」に関する結果

	2017	2018	2019	2020	2021	2022	まとめ
日経Tech Find	温度 加速度 地磁気 センサー 環境 モーション カメラ 人感 照度 気圧	温度 照度 濃度 計測 CO2 分析 室内 気圧 熱	温度 大気 測定 騒音 CO2 照明 LED 反応、水 音圧 気圧 照度 濃度	湿度センサ 土壌 データ 指示 日照 テープ 実測 LED マニュアル 散布 不快	温度 データ 取得 収集 活動 気圧 監視 カメラ 日照	事業 脈拍 細胞 冷風	農業分野
	項目なし	項目なし	日常生活 Solace Power Alex- Lidow アクティビティ 後継 予告	項目なし	高機能 コンセプト 組み合わせ 歩留まり 全体的	スイッチ ディスプレイ マルチ リレー 薄型	システム 全体
	resi- dents radars splin- klers	enter- prise	rate slowly	Atom- osphere medi- cinate outdoor	clean- ness datalink respon- sibility	項目なし	使用用途

図4 ホットワード「湿度」に関する結果

## 4.2 考察

酸素センサと湿度センサに対して記事の収集を行い、自然言語解析を実施したが、「日経 TechFind」「Anews」「Shareresearch」で同じテーマを扱っているにもかかわらず、それぞれの出力結果は異なっており、傾向も媒体ごとに異なっていた。また市場調査では自動車産業が大きなシェアを占めており、同様に、技術者の頭の中では「1.車載(自動車産業)、2.医療、3.産業機器、4.家電、5.モバイル、6.IoT」との回答だったが、今回の解析結果では主に車載以外の項目がピックアップされ、一般的な回答とは異なる結果を得ることが出来た。

## 5 おわりに

本研究では、市場探索の1つの手法として、情報リソースごとに年代ごとの自然言語処理を行い、重要単語をピックアップした。これを時系列に並べることでその変遷が分かるようになった。また「日経 TechFind」「Anews」「Shareresearch」の3つの媒体を同一テーマ(「酸素センサ」「湿度センサ」)で自然言語処理解析することで比較をした。さらに出力結果と、市場調査及び(技術者に対する)事前ヒアリングとの比較を行うことで自動車産業(車載)以外の市場についての見出しを行うことが出来た。これは、(同じテーマであっても媒体ごとに出力結果が異なっており)解析者が多様な見方をすることにつながっていると考えられる。人力調査や一般論では見落としていた情報や市場が見えることになり、これか

## 参考文献

[1]上田紗綾.特許情報の時系列解析結果と売上データを利用した半導体に関する重要技術の抽出方法.(オンライン)(引用日:2023年1月11日.)  
[https://www.anlp.jp/proceedings/annual\\_meeting/2022/pdf\\_dir/G6-5.pdf](https://www.anlp.jp/proceedings/annual_meeting/2022/pdf_dir/G6-5.pdf)  
[2]株式会社グローバルインフォメーション.酸素ガスセンサー市場-成長、動向、COVID-19の影響、予測(2022年-2027年).(オンライン)(引用日:2022年12月28日.)  
<https://www.gii.co.jp/report/moi393345-global-oxygen-gas-sensors-market-growth-trends.html>

らの技術開発の多様化を促し、開発の一助となるだろうと考えている。

今後は、技術者ヒアリングについて今回は1名のみであったので、特定テーマの複数人員に対して事前ヒアリングを行いつつ自然言語処理を組み合わせる手法について検証を行いたい。また、今回は単語による市場探索を行ったが、文脈理解、文章や文脈の単位でクラスタリングできる手法が適切であるとも考えられる。今後はトピックモデルやBERTの適用も検討していく。

[3]株式会社グローバルインフォメーション.湿度センサーの世界市場(2022-2027年).(オンライン)(引用日:2022年12月28日.)  
<https://www.gii.co.jp/report/imarc1071045-humidity-sensor-market-global-industry-trends.html>  
[4]小沼良直.民間企業における研究開発テーマ評価と研究開発者評価.(オンライン)(引用日:2023年1月11日.)  
[https://www.mext.go.jp/component/a\\_menu/science/detail/\\_icsFiles/afieldfile/2010/06/25/1294227\\_8.pdf](https://www.mext.go.jp/component/a_menu/science/detail/_icsFiles/afieldfile/2010/06/25/1294227_8.pdf)  
[5]経済産業省.研究開発に係る無形資産価値の可視化研究会.(オンライン)(引用日:2022年12月28日.)  
[https://www.meti.go.jp/shingikai/economy/mukei\\_shisan/index.html](https://www.meti.go.jp/shingikai/economy/mukei_shisan/index.html)

- [6] 難波英嗣.手順オントロジー構築のための特許請求項の構造解析.(オンライン)(引用日:2023年1月12日.)  
[https://www.japio.or.jp/00yearbook/files/2020book/20\\_3\\_01.pdf](https://www.japio.or.jp/00yearbook/files/2020book/20_3_01.pdf)
- [7] 安藤俊幸.特許調査のためのプログラム事例紹介.(オンライン)(引用日:2023年1月12日.)  
[https://www.jstage.jst.go.jp/article/jkg/70/4/70\\_203/\\_pdf/-char/ja](https://www.jstage.jst.go.jp/article/jkg/70/4/70_203/_pdf/-char/ja)
- [8] 安藤俊幸,桐山勉.分散表現学習を利用した効率的な特許調査.(オンライン)(引用日:2023年1月12日.)  
[https://www.jstage.jst.go.jp/article/infopro/2019/0/2019\\_31/\\_pdf/-char/ja](https://www.jstage.jst.go.jp/article/infopro/2019/0/2019_31/_pdf/-char/ja)
- [9] 安藤俊幸.機械学習による予備検索を考慮した効率的な特許調査.(オンライン)(引用日:2023年1月12日.)  
[https://www.jstage.jst.go.jp/article/infopro/2020/0/2020\\_43/\\_pdf/-char/ja](https://www.jstage.jst.go.jp/article/infopro/2020/0/2020_43/_pdf/-char/ja)
- [10] 特許庁.特許出願技術動向調査.(オンライン)(引用日:2023年1月12日.)  
<https://www.jpo.go.jp/resources/report/gidou-houkoku/tokkyo/index.html>
- [11] 山本雄太.技術コミュニティの成長性を加味した特許価値評価手法の開発.(オンライン)(引用日:2023年1月12日.)  
[https://www.jstage.jst.go.jp/article/pjsai/JSAI2020/0/JSAI2020\\_4K2GS305/\\_pdf/-char/ja](https://www.jstage.jst.go.jp/article/pjsai/JSAI2020/0/JSAI2020_4K2GS305/_pdf/-char/ja)
- [12] 米村崇.特許分析を通じた製薬企業の研究開発と企業価値との関連性の研究.(オンライン)(引用日:2023年1月12日.)  
[https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara\\_id=KO40003001-00002020-3761](https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40003001-00002020-3761)
- [13] 柴田洋輔.特許と論文の複合解析による有望応用分野の予測—印刷技術を例に—(オンライン)(引用日:2023年1月12日.)  
[https://www.jstage.jst.go.jp/article/jkg/69/3/69\\_128/\\_pdf/-char/ja](https://www.jstage.jst.go.jp/article/jkg/69/3/69_128/_pdf/-char/ja)
- [14] 日経 TechFind. トップページ.(オンライン)(引用日:2022年12月28日.)  
<https://xtech.nikkei.com/service/techfind/>
- [15] Anews. トップページ.(オンライン)(引用日:2022年12月28日.)  
<https://anews.stockmark.ai/#/account/login>
- [16] Shareresearch.(オンライン)(引用日:2022年12月28日.)  
<http://demosc.shareresearch.net/sr-c/Login/Login.aspx>
- [17] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean (2013) “Efficient Estimation of Word Representations in Vector Space” . (online) (引用日:2023年1月12日.)  
<https://arxiv.org/pdf/1301.3781.pdf>
- [18] Laurens van der Maaten, Geoffrey Hinton. Visualizing Data using t-SNE(online) (引用日:2022年1月12日.)  
<https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf>
- [19] a b Simoudis, Evangelos; Han, Jiawei; Fayyad, Usama M., eds (1996). “A density-based algorithm for discovering clusters in large spatial databases with noise”. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press. pp. 226–231. ISBN 1-57735-004-9 (引用日:2022年1月12日.)  
<https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf>