

# 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

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

技術開発において、どんな技術が売上に貢献できそうか、という判断は重要である。この目的達成のため、市場に供する事を目的とする技術が記載された特許の解析は非常に重要である。特許は、学術論文に比して発行量が多く、精読してその全容を把握することは常に困難である。そこで、本研究では、特許文献に対して自然言語処理を行い、何が重要技術であるかを抽出可能な手法を提案する。重要かどうかの判断は、対象技術を保有する企業の売上の推移という検証可能な数値と比較することで行った。

具体的には、既知の技術である白色 LED(Light Emitting Diode)の 1990 年代から 2000 年代にかけての大量の特許情報に対して適用し、白色 LED の売上が拡大し始める年の 3 年前に、その基礎となる技術が出願されていたことを明らかにした。

## 1 背景

知的財産は産業上重要な権利であるが、権利の範囲やその権利の所有者の情報などから各企業の将来にわたる技術のトレンドを表すものとして、分析の対象となってきた。近年ではそうした分析の手法として、各統計データを見える化する「知財ランドスケープ」の概念の活用が進んでいる[1]。しかしながら、一度に多数の特徴量を考慮することが難しく、出願特許の個々の詳細や傾向について適切な解析はできない。

そこで、本研究で既存の解析の改善を試みる事とした。そのために自然言語処理を活用し、多次元の特徴量を言語ベクトルとして出願特許から抽出し、そのどれが重要か、を売上の推移との相関があるかどうかで決定する、という手法を取った。

## 2 関連研究

大量の出願特許における解析手法の研究としては、新規性の評価[3][4][5][6]、概要の要約[7]、特許記載

技術の価値判断などがある。この内本研究で対象とする、特許記載技術の時系列解析を用いた価値判断手法としては、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[8]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[9]、論文本文数を用いて評価する手法[10]がある。しかしこれらの手法は、客観的な数値に基づく評価は行っているものの企業全体の評価にとどまっており、技術トピックスのどの部分がそれぞれの数値に貢献しているかまでの分析は行っていない。

これらに対して本研究では売上の推移に着目し、出願特許を年ごとに単語ベクトル化し売上の推移と比較することで、各技術における重要なトピックスを見つけ出すことを提案する。また手法の評価のために、既知の技術である白色 LED について、市場に出始めた 1990 年代から 2000 年代前半にかけての LED に関する特許を調査し、事前の知識を前提とせず白色 LED の市場拡大となった技術を抽出することに成功した。企業の評価だけでなく、売上拡大につながる技術の特定まで行うことで、企業における技術開発の方向性を判断する材料を提供できると考えている。

## 3 提案手法

本章では最初にデータセットの収集方法及び解析手法を示す。

### 3.1 解析対象とするデータ

データは 1986 年から 2021 年に出願された白色 LED 特許を収集した。特許総数は 7552 本ですべて米国特許である。

### 3.2 アルゴリズム

本手法の提案アルゴリズムは Word 2vec における「skip-gram」を基本とする[11]。「skip-gram」の特徴は、中央の単語(ホットワード)から周囲の複数ある

単語を推測するところである。中央の単語(ホットワード)を $w_t$ とした時、その前後に存在する可能性が高い単語 $w_{t-1}$ と $w_{t+1}$ について推測する。これは $w_t$ が与えられた時に、 $w_{t-1}$ と $w_{t+1}$ が同時に起こる条件付き確率 $P(w_{t-1}, w_{t+1} | w_t)$ を計算することで行う。

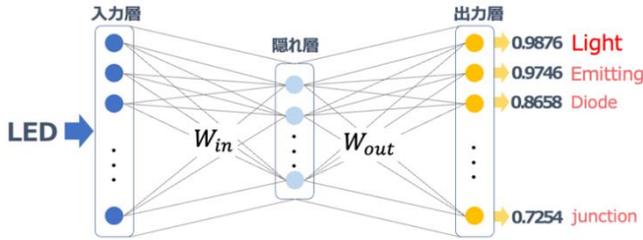


図1 本手法のアルゴリズム

図1にその概要を示す。中央の単語に対して、文章中のすべての単語について条件付き確率を計算する。最終的には最も大きい数値の単語が、中央の単語(ホットワード)に対して最も関連性があると判断する。

モデルは「入力層」と「出力層」、およびNN(ニューラルネットワーク)を構成するために複数の「隠れ層」がある。「入力層」に単語が入力されると、「隠れ層」によって入力側 $W_{in}$ と出力側 $W_{out}$ に対して重みの更新が行われる。例えば、ある自然言語データにおいて「LED」という単語の周辺に「Light」や「Emitting」という単語が出現したとする。この場合、入力層の「LED」に対応した要素に1、それ以外の要素に0が入力される。「Light」、「Emitting」に対応した出力層の出力が1に近づくように $W_{in}$ と $W_{out}$ が更新される。更新式は(1)(2)式のようになる。なお、ここでは $W_{in}$ を「 $W$ 」、 $W_{out}$ を「 $W'$ 」で表記している。 $v$ は1つの単語ベクトルで $all$ は文章中に登場する全ての単語に関する単語ベクトルの集合を意味する。 $W_i$ は $i$ 番目の入力単語の行ベクトルで、 $W'_j$ は $j$ 番目の出力単語の列ベクトル、 $i$ は各ベクトルの $i$ 番目の重み、 $\eta$ は学習率で、0から1の範囲をとる。また $t_j$ は $j$ 番目の出力単語が $i$ 番目の入力単語に対して正解単語であるかどうかの可能性を表している。

$$W_{ij}^{(new)} = W_{ij}^{(old)} - \eta \sum_{v \in all} (p_v - t_v) W'_{iv} \quad (1)$$

$$W'_{ij}^{(new)} = W'_{ij}^{(old)} - \eta (p_j - t_j) W_{ij} \quad (2)$$

ここで $P_j$ は $i$ 番目の入力単語と $j$ 番目の出力単語の共起確率を表しており、式(3)で表される。

$$P_j = \frac{\exp(W_i \cdot W'_j)}{\sum_{v \in all} \exp(W_i \cdot W'_v)} \quad (3)$$

### 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。取得した特許に対して単語ごとに分解する「分かち書き処理」を行う。次に1つの単語に対して1つのベクトルを割り当て、文章中に登場するすべての単語をベクトルで表現する。作成されたベクトルは多次元(本手法では100次元)となっているのでt-SNE(Stochastic Neighbor Embedding)[12]により次元削減を行い人間が理解できる2次元マップを作成する。最後に、作成された2次元マップ上の各点について、近傍にある点同士をDBSCAN(Density-based Spatial Clustering of Application)[13]によってクラスタ化した。それぞれのクラスタがどんな意味を持つのか、については本研究グループで判断した。

### 3.4 LED 特許解析

提案フローの評価のために、すでに既知の技術である白色LEDに関して、売上が急拡大した1990年代にどのような技術トレンドがあったのかを評価した。まず入力データについては、1986年から2021年の米国LED特許を選定し、分析をしたいホットワードとして、白色LEDを最もよくあらず単語である「LED」と「white」を選択した。次に年代ごとに「skip-gram」モデルによる機械学習を実行したのち、年代ごとの2次元分布マップを出力する。最後に出力された分布マップを年代ごとに並べることにより時系列変化を明確にする。実際に出力したマップの例として、1990年の「LED」をホットワードとした出力分布マップを示す。「赤色」グルーピング群は「junction」「hetero」や「un turbulent」といった「異なる物質の安定的な接合について」のクラスタが、

「青色」グルーピング群は「Blue」「reacting」や「semiconductor」といった「青色反応と半導体について」のクラスタが構成された。このようにしてグルーピングされた群に対して群名を決定しこれを全

ての年代に対して行い、時系列に並べる。なお出力分布マップの X 軸と Y 軸は t-SNE により次元圧縮される過程で生成したものであり、無次元である。

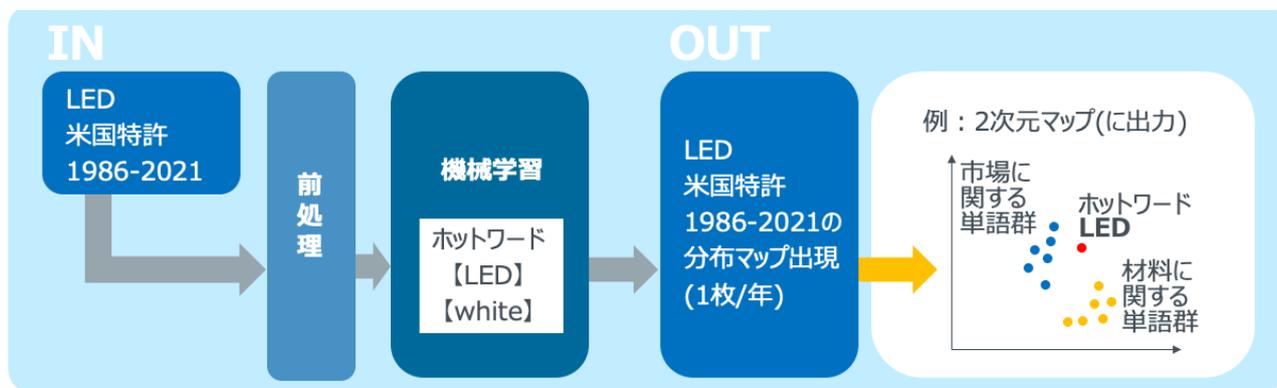


図 3: 本手法の処理フロー(LED 米国特許解析 1986-2021)

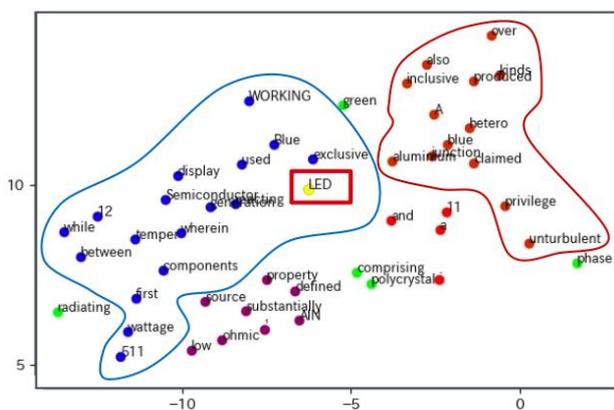


図 4: ホットワード LED に関する出力分布マップ

## 4 結果

ここでは「3.4 出力分布マップ」で出力された年代ごとのマップの解析結果について述べる。下記マップは白色 LED の市場を創生した日亜化学工業の業績推移[14]である。白色 LED の売上は 1998 年から生じており、その後の売上拡大の起点であることがわかる。よってこの年代の前後で何らかのホットワードが取り出せれば、それに関する技術が売上に貢献した可能性が高いと考えた。

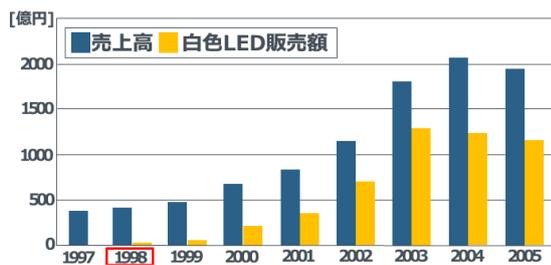


図 5: 日亜化学工業の業績推移

### 4.1 ホットワード「LED」によるクラスタ

解析を行い出力されたマップを時系列に並べた結果、1990 年は「青色 LED」について、1995 年から 1998 年の間には「白色 LED」「光」「RGB マトリックス」について、2000 年には「照明の生産について」のクラスタの存在が確認できた。

マップによる解析結果と白色 LED の売上げを併記した結果を図 6 に示す。1998 年が白色 LED の売上拡大の起点であるが、本手法では 3 年前の 1995 年は「白色 LED」がコア技術であったことがわかる。

つまり、市場を形成する数年前に有効な「白色 LED 特許」が出願されており、この特許群は市場獲得に先行している。また白色 LED という技術が市場を獲得するまでに 3 年間のタイムラグが発生していることがわかった。

### 4.2 ホットワード「white」によるクラスタ

次に「white」をホットワードとして解析を行った。「LED」同様にマップと売上げを併記したものを図 7 に示す。「white」の場合は 1994 年に「アプリケーション」についてのクラスタ、1997 年から 2000 年の間では「光を効率的に役立たせることについて」クラスタが出現していることが確認でき、「LED」同様売上が生じる数年前から特許化が始まっていることが判明した。「white」の変遷についてまとめると「アプリケーション」「光の効率」「生産」「LED、バックライト、ヘッドライト」「開発保証」「一般化」といった状況が解析の結果得られ、こうした技術が量産に貢献していたものと推測される。

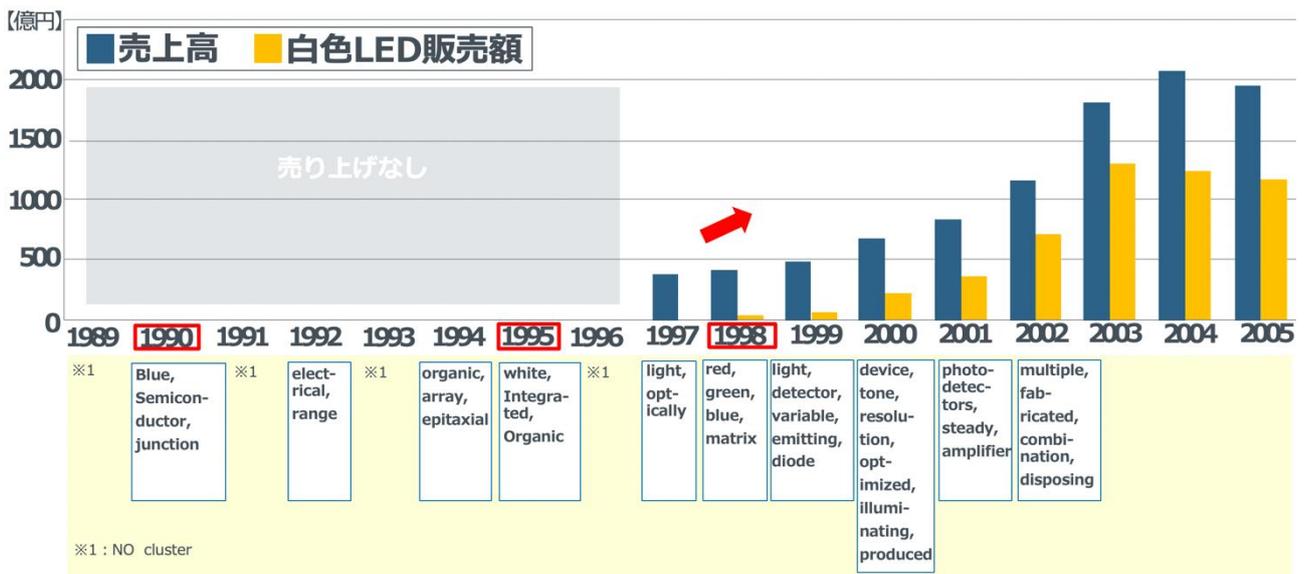


図6 日亜化学工業の業績推移にホットワード「LED」を重ね合わせたもの

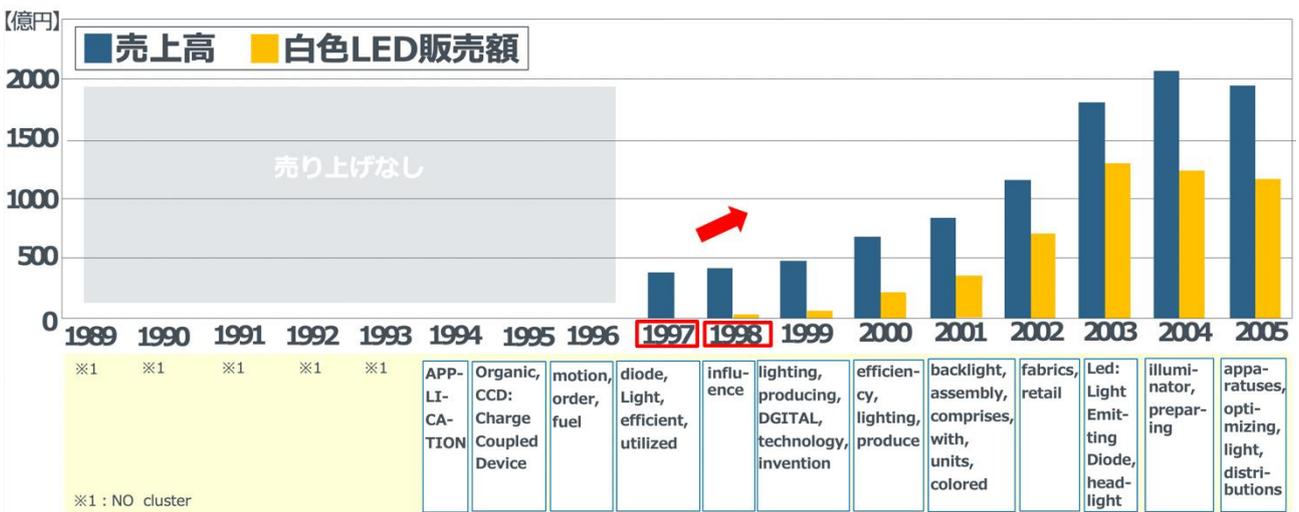


図7 日亜化学工業の業績推移にホットワード「white」を重ね合わせたもの

## 5 おわりに

本研究では、特許情報の解析手法の1つとして、ホットワードに関する重要な単語群を年代別に取り出し、それらを実際にも実証可能な数値と比較することで、売上拡大につながった重要な技術を取り出すことに成功した。本手法は、解析を担当する者が解析分野に関する詳細な知識を事前に知っておく必要がないため、特許以外にも活用が可能であると考えられる。

一方で本手法は特許情報が多数公開されている技術分野には有用であるが、各企業や大学において特許化されていない分野、ノウハウといった技術分野

については、適用することが難しい。また一般的にも使われる単語がホットワードや重要な単語群となる技術については、ホットワードの意味づけが難しく、やはり解析が難しい。そうした技術分野では、単語ではなく文節や文章の単位でクラスタリングできる手法がより適切であり、今後はトピックモデルやBERTへの適用も検討していく。

## 参考文献

- [1]特許庁. 知的人財スキル標準 (Version2.0) [https://www.jpo.go.jp/sesaku/kigyo\\_chiza/chizai\\_skill\\_ver\\_2\\_0.htm](https://www.jpo.go.jp/sesaku/kigyo_chiza/chizai_skill_ver_2_0.htm)
- [2]Patent Result. 企業の成長性がひと目で分かる「権利者スコアマップ 時系列分析」を開発.(オンライン)(引用日:2021年12月13日.)  
<https://www.patentresult.co.jp/news/2011/02/jikeretsu.html>
- [3] 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:2021年12月13日.)  
[http://nlp.indsys.chuo-u.ac.jp/pdf/2020/nanba\\_ifat202003.pdf](http://nlp.indsys.chuo-u.ac.jp/pdf/2020/nanba_ifat202003.pdf)
- [4] 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:2021年12月13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)
- [5] 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:2021年12月13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)
- [6] 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:2021年12月13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)
- [7]特許庁. 特許出願技術動向調査.(オンライン)(引用日:2021年12月13日.)  
<https://www.jpo.go.jp/resources/report/gidou-houkoku/tokkyo/index.html>
- [8]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:2021年12月13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)
- [9]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:2021年12月13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)
- [10]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:2021年12月13日.)
- [11] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean (2013) “Efficient Estimation of Word Representations in Vector Space”. (online) (引用日:2022年1月12日.)
- [12]Laurens van der Maaten, Geoffrey Hinton. Visualizing Data using t-SNE(online) (引用日:2022年1月12日.)
- [13] 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.)
- [14]宮地電機株式会社 照明・LED担当室.LEDの市場(オンライン)(引用日:2021年12月12日.)  
[http://www.eco-glass.com/ibox/data/07-11-13\\_LED-Communication-06.pdf](http://www.eco-glass.com/ibox/data/07-11-13_LED-Communication-06.pdf)