

業績要因・結果を用いた特許の価値判定のための関連性推定

三田英理 酒井浩之 永並健吾

成蹊大学 理工学部 情報科学科

us182054@cc.seikei.ac.jp {h-sakai, kengo-enami}@st.seikei.ac.jp

概要

決算短信の業績要因と特許を関連付け、業績要因のセグメントの業績結果から特許の価値判定のための関連性推定を行う。具体的には、特許と業績要因を類似度推定により関連付けを行う。そこで取得した業績要因を基に、対応する業績結果を取得し、最終的に特許・業績要因・業績結果の3つを関連付ける。3つを関連付けることにより、特許が属するセグメントの業績結果から、出願企業における特許の価値を判定することが可能となる。評価実験では、無作為に選んだ100件の特許に判定を行い、76件が関連付けに成功した。

1. はじめに

企業にとって、競合他社の特許を検索し、先行技術がないか、自社の製品が他社の特許を侵害していないかを調べることは重要である。そのため、知財戦略や研究開発戦略の立案の際、特許文書を対象とした技術調査を行う必要がある。しかし、年間に出願される特許の数は膨大であり、1つの企業が出願する特許数も多い。そのため、たとえ業種を絞ったとしても、全ての特許を対象として調査することは多大なコストがかかり、できるだけその企業にとって価値のある重要な特許のみを調査対象としたい。ここで、企業にとって重要な特許とは、その企業が展開している事業（セグメント）の中でも主力となる事業に関連している特許である。しかし、手作業で競合他社の特許の価値を判定するには、まず、多くの文章量である特許文書からその特許に対応する事業を探し出さなければならない。そして、その事業が企業にどの程度影響を与える可能性があるかを計算する。これは非常に手間のかかる作業であるため、効率的な支援技術が必要である。

そこで本研究では、特許と出願企業の決算短信における業績要因[1]とを関連付け、業績要因が属する

事業セグメントの業績結果から、その特許の関連性推定を行う手法を提案する。そして、特許を出願した企業においてどれだけの価値のあるものであるかを推定する。すなわち、特許の対応する事業セグメントの売上げがその企業全体の売上げの割合を占めるかで、その特許の価値を判定する。

特許分析に関する関連研究として、佐々木らは、特許文献中の重要語を用いたFタームの自動付与を行っている[2]。酒井らは、技術調査に必要な情報を「ができる。」と「が可能である。」といった表現を、手がかり表現の種として与えることで、特許から自動的に抽出する手法を提案している[3]。これらの研究に対して本研究では、特許の出願企業における価値を判定するための情報を、その企業の決算短信の情報を用いて推定するためのものであり、特許の価値判定のための情報抽出、および関連性推定を行う点が異なる。

2. 提案手法

本手法の概要を以下に示す。

- Step1: 特許文書に付与されているIPCタグを用い、出願企業が発行している決算短信における、その特許に最も関連する業績要因文を推定する。
- Step2: Step 1で特許と関連があると推定された業績要因文に対応する業績結果文を推定する。
- Step3: 推定された業績結果文から、その業績結果の対象としている事業セグメントにおける売上高や営業利益を抽出する。
- Step4: その企業全体の売上高や営業利益と、Step 3で抽出された、特許に関連している事業セグメントの売上高や営業利益を比べ、全体における割合からその企業における特許の価値を判定する。

本研究にて使用する特許文書は2017年に出願された特許のうち、上場企業のものを使用する。決算短信は2015年から2020年のデータを使用する。

2.1 特許文書と業績要因文との関連付け

2.1.1 特許文書の分散表現

特許文書と決算短信の業績要因文との関連付けのための類似度計算の準備として、word2vec[4]による特許文書の分散表現を作成する。本研究では、特許文書に付与されている国際特許分類(IPC)から名詞を取得し、分散表現を作成する。なお、word2vecの分散表現モデルは上場企業の有価証券報告書

(30587 個) から作成したものを使用し、次元数は300とした。分散表現モデルの作成に有価証券報告書を用いた理由は、大量の文書が容易に入手できることと、特許と決算短信で同じ分散表現モデルを使用する必要があるため、特許よりも決算短信に近い文体である有価証券報告書を用いた。

IPCは、特許の分野に相当であると認められる全知識体系を、セクションからサブグループまでの階層構造で表現している。したがって、特許文書中のIPCから上の階層のもので構成される名詞から、特許の分散表現を作成する。この名詞の分散表現ベクトルの平均を求めることで、特許の分散表現を得た。名詞の抽出には、形態素解析器 MeCab を使用した。

2.1.2 業績要因文の分散表現

企業ごとの業績要因文の分散表現を作成する。ここで、決算短信からの業績要因文は、酒井らの手法[1]を使用して抽出した。業績要因文ごとに複合名詞を取得し、2.1.1.節と同様に word2vec を用いて分散表現を得る。word2vec のモデルは、2.1.1 節と同じ、有価証券報告書から作成したものである。業績要因文の分散表現も、その文に含まれる名詞の分散表現ベクトルの平均を求めることで得た。

2.1.3 特許と業績要因文との関連性推定

特許文書から作成した分散表現 A と業績要因文から作成した分散表現 B を比較して、特許と業績要因文との類似度を計算する。

類似度としてコサイン類似度を求めることで、特許文書に対して関連性の高い業績要因文を得ることができる。ここで、求めた類似度のうち値が大きかった上位 5 文の業績要因文を特許文書との関連性の

高い業績要因文とした。

2.2 業績要因文に対応する業績結果の抽出

2.1 節で特許文書と関連付けを行った業績要因文に対して、対応する業績結果の取得を行う。ここで、業績結果として、以下の 2 つの条件をともに満たす文を抽出する。

- “売上高” “営業利益” “純利益” のいずれかを含む
- “万円” “億円” のいずれかを含む

業績結果は業績要因文と同じ文、もしくは直後の文に含まれているという特徴を利用した。例として、「島津製作所」の「医用機器事業国内市場」に関する業績要因文と、その後ろ 4 文を以下に示す。なお、1 文目が業績要因文である。

- I I . 医用機器事業国内市場は、内外メーカーの競合激化による価格低下が進むなか、早期ガン検診に適した PET (陽電子放射断層撮影装置)、FPD (フラット パネル ディテクタ) 搭載の循環器 X 線検査システムなどの新製品を中心に好調に推移しました
- 海外市場につきましては、低調でありました
- この結果、当事業の売上高は 4 4 2 億 9 千万円 (前期比 6.8%増)、営業利益は 2 2 億 9 千 9 百万円 (同 7.1.7%増) となりました
- I I I . 航空・産業機器事業航空機器は、防衛予算の抑制や民間航空機需要の低迷など厳しい事業環境が続くなか、国内市場、海外市場ともに低調に推移しました
- 産業機器は、半導体・IT 関連設備投資の増加に伴い、半導体製造装置向けターボ分子ポンプ、太陽電池用成膜装置、ガラスワインダーなどの需要が拡大し、また、産業車両や建設機械向け小型ギアポンプなど油圧機器も好調に推移しました

例において、業績要因文から 2 文目 (先頭からは 3 文目) に業績要因文に対応する業績結果文がある。このように、決算短信では業績結果を含む文は業績要因文の周辺に記載されている。ほとんどの業績要因・業績結果はこの規則に沿って記載されているため、これを利用した抽出を行う。具体的な抽出手順は以下の通りである。

- Step1: 類似度が最も高かった業績要因文を含む後ろ 5 文を抽出して上記のキーワード検索を行う。
- Step2: キーワードに一致する語があった場合は、その文章を業績結果として抽出する。

⁴<https://www.jpo.go.jp/system/patent/gaiyo/bunrui/ipc/ipc8wk.html>

Step3: キーワードに一致する語がなかった場合は、2番目に類似度が高かった業績要因文に対してキーワード検索を行う。

以上の手順を5番目に類似度が高かった業績要因文まで繰り返し行うことで、業績要因文に対応する業績結果文の抽出を行った。

3. 評価

まず、分散表現を用いた類似度計算により取得した業績要因文と特許に対して、正しく関連付けが行われているかどうかの評価を行う。特許に対する最も高い類似度の業績要因文を記載した例を以下に示す。

■企業名： 積水化成工業株式会社

・特許： 容器

・類似度： 0.80565

・業績要因文：自動車部材・産業包装材関連では、「ピオセラン」（ポリスチレン・ポリオレフィン複合樹脂発泡体）は、世界同時不況後の販売回復に加え、中国での需要拡大などにより前期から売上が増加しました

表1に最も高い値の類似度の分布を示す。特許文書に対して最も高い値となる業績要因文の類似度は0.7から0.8を中心に見られ、97.9%が0.6以上となった。

表1 特許に対する類似度の分布

類似度	~0.5	0.5~0.6	0.6~0.7	0.7~0.8	0.8~
合計数	62	1303	9260	36046	18772

続いて、取得した業績結果に対する評価を行う。

2.1節で抽出した業績要因文に対してどの程度の割合で業績結果が取得できているかに着目する。業績要因文65,443件に対して本手法を実行した結果、取得した業績結果は55,290件であり、84.5%となった。

最後に、特許文書と業績要因、業績結果の3つが正しく関連付けできているかについて評価を行う。この評価は取得した特許、業績要因、業績結果の対応から100件を無作為に選び、手作業で評価を行った。100件のうち、特許と業績結果の関連付けができたものが76件となった。例として本手法にて関連性ありと推定した2件の特許、業績要因、業績結

果の対応を以下に示す。

■企業名： アサヒグループホールディングス

・特許タイトル： 血中アセトアルデヒド低減剤

・業績要因文：国内食品事業「アサヒフードアンドヘルスケア株式会社」については、指定医薬部外品『エビオス錠』やサプリメント『ディアナチュラ』などの主要商品に加え、ダイエットサポート食品『スリムアップスリム』や栄養調整食品『1本満足バー』なども好調に推移したことなどにより、売上が前年同期より伸長しました

・業績結果文：以上の結果、食品事業の売上高は、東日本大震災の影響を受けたものの、グループ各社が主力商品のブランド強化に取り組んだことにより、前年同期比 4.5%増の 476億3千3百万円となりました

■企業名： 出光興産

・特許タイトル： グリース組成物

・業績要因文：また、従来の結晶性ポリプロピレン樹脂と比べて大幅に融点が低く軟質特性を有する機能性軟質ポリプロピレン（商品名：エルモーデュ®）については、従来からの衛生材向け接着基剤、不織布の改質材などに加え新たな用途開拓に国内外で取り組みました

・業績結果文：以上の結果、石油化学製品セグメントの売上高は、ナフサ価格が上昇したことなどにより5,007億円（前期比+8.6%）となりました

4. 考察

まず、特許文書と業績要因の関連付けに関する考察をする。特許文書65,443件のうち、84%が0.7以上の類似度となった。特許文書のIPC分類の名詞における分散表現の平均と、業績要因文の複合名詞における分散表現の平均で類似度を計算したことを踏まえると、高い値となった。他にも特許文書の請求項に含まれる全ての名詞による分散表現を用いた手法などが考えられたが、その場合、より専門的な単語が多く含まれる一方で、ノイズとなる単語もまた多く含まれてしまうため、精度が落ちると考える。

続いて、業績要因文に対応する業績結果の取得率に関して、業績要因文を含む後ろ5文においてキーワード検索にて関連付けを行った。その結果、取得率は84.5%となり良好な値となった。正しく業績結果が抽出できなかった場合の傾向として、業績要因文は正しく抽出できているが、5行以内に業績結果が記載されている文が存在しない場合が多くみられた。しかし、このような状況に対応するため、抽出対象の文数を増やしてしまうと、異なる事業セグメントの業績結果を抽出する可能性が高くなる。そこ

で、「売上高」等のキーワード検索だけでなく、業績要因のセグメントと異なるセグメントになり得る語が見つかった場合に、抽出することがないような処理を追加することで改善できると考える。

最後に、特許文書・業績要因・業績結果の3つの関連付けに対して考察を行う。人手で評価を行った結果、100件のうち76件が成功していた。つまり、業績結果が取得できたものに関しては比較的高い精度で関連付けができていていると考える。

5. 特許の価値判定の試み

特許、業績要因、業績結果の3つの情報が関連付けできたことにより、その特許文書の出願企業における重要性を判断することが可能となる。例として、実際に取得した業績結果と出願企業の当期間の全体売上高の比較を行うことで、特許の価値判定を試みる。

■日立製作所：

・特許タイトル 半導体検査装置、及び荷電粒子線を用いた検査方法

・当期間における全体売上高 90,410 億円
・業績結果における売上高 10,143 億円 (11.2%)

■リョービ：

・特許タイトル 真空ダイカスト装置および真空ダイカスト法
・当期間における全体の売上高 170,973 百万円
・業績結果における売上高 104,094 百万円 (60.8%)

このことから、リョービの「真空ダイカスト装置および真空ダイカスト法」の特許に関する事業の売上高は企業全体と比較して60.8%であるため、大きな影響を与える可能性を持つ特許であると推定できる。実際に、リョービは独立系ダイカスト専業首位の企業であり、そのような企業にとって「真空ダイカスト装置および真空ダイカスト法」に関する特許は大きな価値があると考えられる。

対して、日立製作所の「半導体検査装置、及び荷電粒子線を用いた検査方法」の特許に関する事業の売上高は企業全体と比較して11.2%であり、あまり大きな影響はない。日立製作所は様々な事業を行っており、その中で「半導体検査装置」が占める割

合は大きくなく、従って、当該特許は日立製作所にとって影響は少ない。

以上のように、業績要因・業績結果を用いて特許に関連する事業を推定することで、その特許が企業に与える影響力を推測することが可能となる。

6. むすび

本研究では、特許の出願企業における決算短信の業績要因・業績結果を用いて、特許がその企業にとってどの程度重要であるか推定するための情報を抽出する手法を提案した。特許文書の分散表現と業績要因文の分散表現により類似度計算を行うことで関連性を推定し、さらに、業績要因文に対応する業績結果文を取得することで、特許の価値を判定するための情報を得ることができた。

評価の結果、業績要因に対して取得できた業績結果は84.5%、100件の特許に対して、業績要因、業績結果の関連付けができたものが76件となり、比較的良好な結果を得ることができた。

今後の課題として、今回の手法では、1つの特許が複数事業に影響を与える可能性についてのアプローチができていないため、複数事業への対応が必要であると考えられる。

参考文献

- [1] 酒井 浩之, 松下 和暉, 北島 良三: “学習データの自動生成による決算短信からの業績要因文の抽出”, 日本知能情報フェスティバル学会誌, vol. 31, no. 2, pp. 653-661, 2019
- [2] 佐々木 深, 綱川 隆司, 西田 昌史, 西村 雅史: “特許文献中の重要語を用いた F ターム自動付与”, 言語処理学会 第23 回年次大会発表論文集, pp.450-453, 2017
- [3] 酒井 浩之, 野中 尋史, 増山 繁: “特許明細書からの技術課題情報の抽出”, 人工知能学会論文誌, vol.24, no.6, pp.531-540, 2009.
- [4] Yoav Goldberg, Omer Levy: “word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method”, arxiv 1402.3722, 2014.