

# 特許情報収集による知的財産創造のための 発見的価値創造の手法の開発

富山県立大学大学院工学部 情報システム工学専攻  
1855005 小野田成晃

指導教員：奥原浩之

## 1 はじめに

ICT 分野の発達により、民間団体や政府機関のデータを ICT 化することの重要性が増している。総務省では、オープンデータ戦略の推進と題して、行政の透明性・信頼性の往生、国民参加、官民協働の推進、経済の活性化・行政の効率化が三位一体で進むことを目的として行われている [1]。

例えば、災害関連情報として震源や震度に対するデータベースとその API を公開する試みが行われている [2]。そのうちの一つとして特許情報プラットフォーム<sup>1</sup>がある、そこでは日本の特許庁に提出された特許や実用新案等が掲載されており、Web サイト上で特許をキーワード検索することで特許利用の効率化を図っている。

しかし、これらのオープンデータは人手で少数の特許事例を調べるのには必要充分であるが、ビックデータとして特許全体の分析を行いたい場合は整理されているとはい難い、例えば、データの保存形式が PDF 担っている場合や、被引用特許の件数が掲載されていない等の問題点がある。

そこで、本研究では特許データのような複雑なパラメータの組み合わせを統計的・数理計画的な側面からの評価するため、日本語の特許データに対して DEA(Data Envelopment Analysis) を用いる。以上の試みにより、価値のある特許のモデル構築手法を提案する。

## 2 DEA

### 2.1 DEA

DEA とは Charnes らが提唱した経営分析手法 [3] であり、絶対尺度の存在しない企業等の組織評価において、DEA が相対尺度を提示する手法として用いられる [4]。具体的には、分析対象となる組織を DMU(Decision Making Unit) に対して産出と投入の比から各 DMU の生産性を測定する手法である。支店  $k$  の効率性の線形計画として定式化したすると以下の式で表せる。

$$\begin{aligned} \max \quad \theta_k &= \frac{\sum_{n=1}^N v_{kn} y_{kn}}{\sum_{m=1}^N u_{km} x_{kn}} & (1) \\ \text{subject to} \quad & \frac{\sum_{n=1}^N v_{kn} y_{sn}}{\sum_{m=1}^N u_{km} x_{sn}} \leq 1 \quad (s = 1, 2, 3, LK) \\ & u_{km} \geq 0 \quad (m = 1, 2, 3, LM) \\ & v_{kn} \geq 0 \quad (n = 1, 2, 3, LN) \end{aligned}$$

ここで  $x, y$  は入力、出力を表し、 $u, v$  はそれぞれ指標に対する重みとし  $s$  は評価対象の DMU の番号とする。この式では  $k$  の効率性は仮想出力/仮想入力の式で、 $k$  にとって最大の効率値が得られるように重み付けを行う。つまり、自分に対して最も有利に評価するのが DEA の肝である。

### 2.2 CCR モデル

式 (1) を線形計画問題として解くために分母または分子を 1 にする手法を CCR モデルと呼ぶ。分母・分子どちらを 1 にするかは入力と出力のどちらを基準とするかで変わる。今回は入力指向として仮想入力を 1 として式 (2) のように定式化した。このモデルでは、効率性の最大化は仮想出力の最大化に等しくなるので良い出力を得られる DMU を評価できる。

$$\begin{aligned} \max \quad \theta_k &= \frac{\sum_{n=1}^N v_{kn} y_{kn}}{\sum_{m=1}^N u_{km} x_{kn}} & (2) \\ \text{subject to} \quad & \sum_{m=1}^N u_{km} x_{sn} - \sum_{n=1}^N v_{kn} y_{sn} \geq 0 \\ & \sum_{m=1}^N u_{km} x_{sn} = 1 \quad (s = 1, 2, 3, LK) \\ & u_{km} \geq 0 \quad (m = 1, 2, 3, LM) \\ & v_{kn} \geq 0 \quad (n = 1, 2, 3, LN) \end{aligned}$$

式 (2) に対して線形計画問題を解くことで各 DMU の効率値を出すことができる。DEA の概念図として図 1 を示す。図中の効率的フロンティアとは DEA により示された効率的かどうかの境界であり、ここに近ければ近いほど DMU は効率的であるといえ、重なっている DMU は効率値が最大の 1 を取るものとなる。

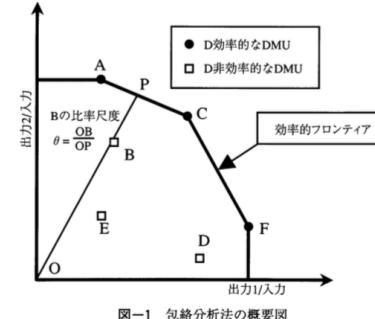


図 1 DEA の概念図

## 3 公開特許データからの収集

### 3.1 収集方法

今回特許の定量的な分析をするためのリソースとして、Google Patent を用いた。Google Patent では、世界各国の特許データが html 形式で表示されていて、PDF などの非構造データに比べてデータ収集しやすい利点がある。

また、Google Patent は API が公開されていたが 2015 年をもって終了したため、独自にデータを収集することにした。まず、必要なデータとして、1. 特許 ID, 2. 発明者, 3. タイトル, 4. 承認日, 5. 引用特許数, 6. 被引用特許数, 7. 本文に含まれる特許内の単語とその頻度とした。また、本研究では津村らと同様、特許に含まれている単語を分析対象としているため、文章から抽出する素性は名詞のみとした [5]。なお収集特許は日本語で提出されたものを対象とした。

### 3.2 データベースの構築

前項で収集したデータを分析するためにデータベースを構築する必要がある。収集データは 7 種類あるが、そのうち単語に関しては各特許に対して抽出できる種類数が異なるため、スケーラビリティのある NoSQL を用いた。また、収集した全特許に対する全単語種を分析で用いるため、別途全単語データベースも構築した。

## 4 シミュレーション結果ならびに考察

### 4.1 DEA 手法の適用

DEA 手法の特許データに対しての適用としては、Hyeonju らの研究がある [6]。Hyeonju らは特許の単語データから新技術の適用領域を探るために用いていたが、本研究では入力を単語の頻度行列に絞ることで、価値のある特許の算出とともに、価値のある特許に含まれる単語の予測も目的の一つとして行う。これにより、特許というドメインでのコアな単語を浮き彫りにできると考えられる。次に具体的な定式化を示す。

本研究では DEA における仮想入力項目を特許に含まれる単語とその頻度とし、仮想出力を引用件数、被引用件数とする。これは後藤らの研究から特許の価値で重要な素性として被引用特許数と引用特許数が挙げられたためである [7]。通常 DEA では仮想入力は小さく、仮想出力が大きいほど効率的であるとするが、本研究のもとでは仮想入力が単語のため、このルールが適用可能かどうかは不明である。そのため 2 つのルールを設け DEA を適用した。

ルール 1 特許  $i$  中の単語  $w_i$  の出現頻度  $n_i$  の逆数を入力とし、出現が 0 であれば入力を定数  $Z$  とおく

ルール 2 そのまま単語の頻度分布を入力とする

ルール 1 は含まれている単語数が少ない割に仮想出力が大きければ良いという仮設に基いて行い、価値の高い特許モデルを生成することを目

<sup>1</sup><https://www.j-platpat.ipit.go.jp/web/all/top/BTmTopPage>

的とする。ルール 2 では良い特許での単語分布を予測するために用いる。ただしルール 1 と 2 の効率である DMU を比較しルール 1 を教師として、ルール 2 から生成されるモデルの精度検証を行う。また、入力/出力の概念表を表 1 に示す。

## 4.2 実験結果

実験結果を以下のとおりとなる。

## 5 おわりに

本研究では特許の定量的評価指標の構築とともに出現単語の分布を分析することを提案した。これらの頻度分布を元に技術分野のトレンドの分析や特許の価値を含んだ特許のテンソル化を行えた。以上より、特許データを始めとした様々な変数が入り混じった複合データの分析やマルチモーダルな機械学習の一助となると考えられる。

今後の課題としては、特許分野における専門用語のステミングや NMF による次元縮約で入力次元を減らし、計算量を落とすことがある。

## 参考文献

- [1] 総務省, ”オープンデータ戦略の推進”, [http://www.soumu.go.jp/menu\\_seisaku/ictseisaku/ictriyou/opendata/](http://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/opendata/), Accessed: May 5, 2018.
- [2] NTT コミュニケーションズ, ”情報流通連携基盤の防災・災害情報における実証にかかる請負実地報告書”, 第 1.0 版, 2013.
- [3] Charnes, A., Cooper, W.W., Rhodes, E., ”Measuring the Efficiency of Decision Making Units”, *European Journal of Operational Research*, Vol. 2, , pp. 429-444, 1978.
- [4] 末吉俊幸, ””DEA 経営効率分析法”, 朝倉書店, 2001.
- [5] Hyonju Seol, Sungjoo Lee, Chulhyun Kim, ”Identifying new business areas using patent information: A DEA and text mining approach”, *Expert Systems with Applications*, vol. 38, pp. 2933-2941, 2011.
- [6] 津村拓海, 斎藤史哲, 石津昌平, ”ランダムフォレストを用いた特許に関する文書データからの技術適用領域に関する知識抽出”, 日本経営工学会論文誌, vol. 68, No. 3, pp. 161-170, 2017.
- [7] 後藤晃, 玄場公規, 鈴木潤, 玉田俊平太, ”重要特許の判別指標”, *RIETI Discussion Paper Series*, 06-J-018, 2018.