

# 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

奥原研究室 M1 小野田成晃

# 研究目的

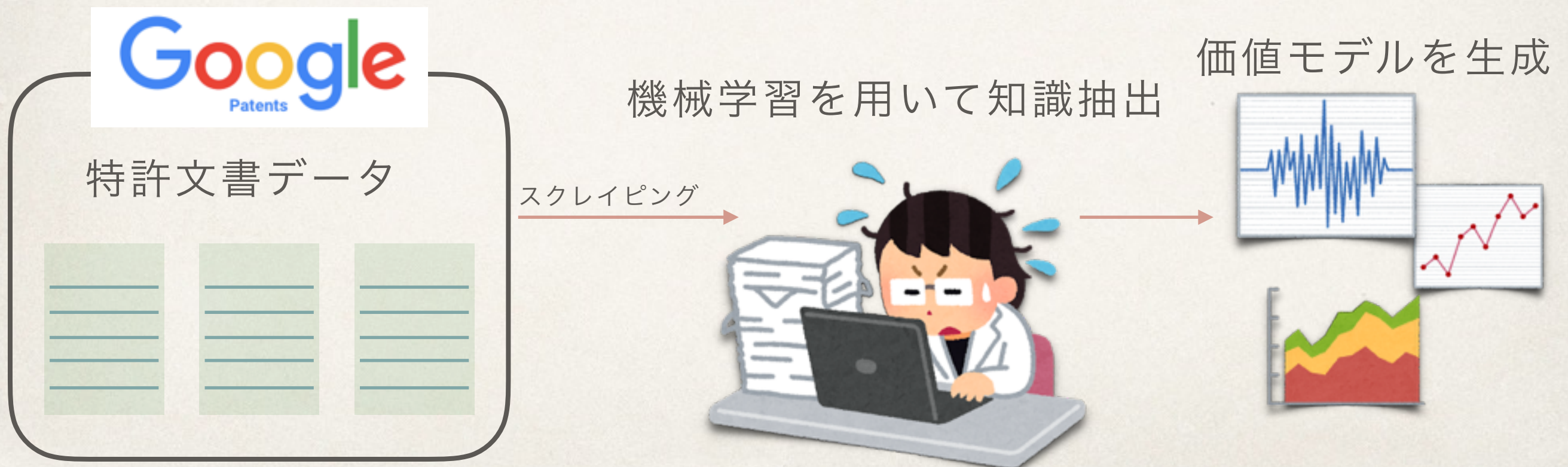
技術情報が多様化・複雑化

- 技術者間での技術傾向に関する情報の共有が**難化**  
そのためにはパテントマップが必要となる

そこで本研究では**WEB**上の公開データから**DEA**を用いて特許価値モデルを作成する手法を提案する

## 研究概要

研究は以下のプロセスで行われる





# 機械学習によるパテントマップ作成

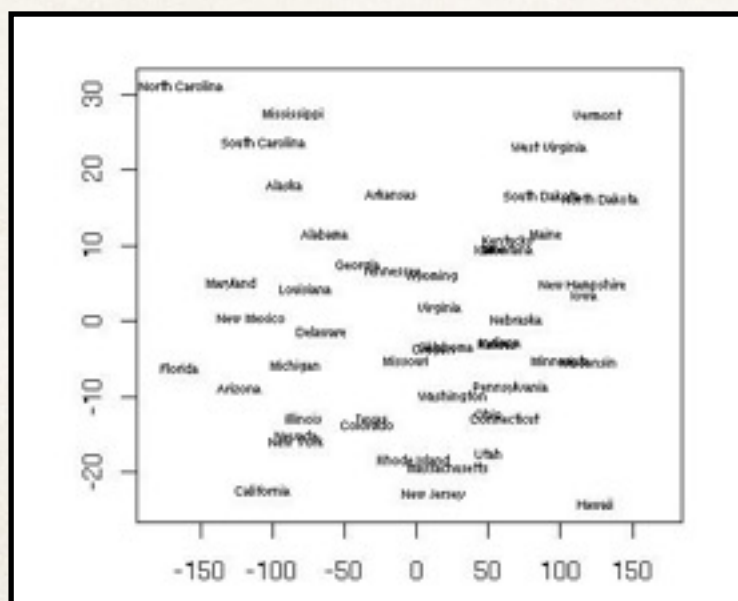
従来：文書間の類似度や決定木によって作成

本研究：文書だけでなく特許価値を考慮

## 一>マルチモーダルな特許価値モデルを生成できる？

# 從來

MINAMOTO,0,70,60,90,70,60,80,100,110,80  
KASAIRINKAI,70,0,60,80,50,40,20,70,80,60  
TONERI,60,60,0,70,50,40,70,80,90,90  
HIKARIGAOKA,90,80,70,0,50,60,70,70,80,100  
YOYOGI,70,50,50,50,0,40,50,40,40,60  
UENO,60,40,40,60,40,0,40,70,80,60  
YUMENOSHIMA,80,20,70,70,50,40,0,70,80,60  
KOMAZAWA,100,70,80,70,40,70,70,0,40,90  
KINUTA,110,80,90,80,40,80,80,40,0,100  
SHINOZAKI,80,60,90,100,60,60,60,90,100,0



## 本研究

DMU	効率値	参観集合をあらわすウェイト											
		1A	1B	1C	1D	1E	1F	1G	1H	1I	1J	1K	1L
A	1	1	0	0	0	0	0	0	0	0	0	0	0
B	0.674	0	0	0	0	0	0	0.404	0	0	0.124	0.054	0
C	0.943	0	0	0	0	0	0	0.889	0	0	0.21	0.113	0
D	0.885	1	0	0	0	0	0	0	0	0	0	0.265	0
E	0.331	0	0	0	0	0	0	0.097	0	0	0.38	0.256	0
F	0.757	0	0	0	0	0	0	0.631	0	0	0	0	0
G	1	0	0	0	0	0	0	1	0	0	0	0	0
H	0.755	0	0	0	0	0	0	0.789	0	0	0.715	0	0
I	0.638	0	0	0	0	0	0	0.276	0	0	0.184	0.368	0
J	1	0	0	0	0	0	0	0	0	0	1	0	0
K	1	0	0	0	0	0	0	0	0	0	0	1	0
L	0.556	0	0	0	0	0	0	0.103	0	0	0.176	0.956	0

0でない値を持つ事業体の集合が事業体Eの参照集合





# データ収集



特許情報プラットフォームは整理されておらず、時間ごとのアクセス制限を設けていたためビッグデータ収集の基盤として不適切

そこで

よく整理されており、マークアップにより構造化されたGoogle Patentを対象としてデータ収集を行った

# ここでデータ収集デモ

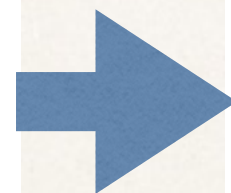
---



# DEA

- ❖ DEAとは各DMUに対して利益が最大となるように重み付け評価をすることで多次元パラメータの効率性を求めるデータ解析手法

$$\left. \begin{array}{l} \max_{u,v} \frac{u^t y_o}{v^t x_o} \\ \text{subject to } \frac{u^t Y}{v^t X} \leq 1 \\ u \geq 0 \\ v \geq 0 \end{array} \right\} \quad (1)$$



$$\left. \begin{array}{l} \max_u \frac{u^t y_o}{v^t x_o} \\ \text{subject to } \frac{u^t Y}{v^t X} \leq 1 \\ -v^t X + u^t Y \leq 0 \\ u \geq 0 \\ v \geq 0 \end{array} \right\} \quad (2)$$

- ❖ 上式のように $v^t x_o$ を1と仮定して線形計画問題に持ち込む手法CCRを用いる

# 本研究におけるDEA

## 2つのルールの結果を比較・検証

- ❖ ルール 1 : 特許 $i$ 中の単語 $w_{\{k\}}$ 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

➡ 単語数が少ない割に仮想出力が大きければ良いDMUという仮説に基づく

- ❖ ルール 2 : そのまま単語の頻度分布を入力とする

➡ 良い特許における単語の頻度分布が予測可能?



# DEA入力概念表

入力

出力

	W1	W2	...	Wk	引用特許	被引用特許
P1	6	0		3	1	0
P2	2	4		0	3	4
⋮						
Pi	1	2		1	1	1



# デモ第二弾

---



# おわりに

---

- ❖ 生成された分布を類似度順にならべることで特許分野におけるトレンド分析が可能
- ❖ また特許の価値を含んだ数値データができるので今後の新しい知財創出の大きな助けとなる
- ❖ 単語のステミング問題が深刻なため解決策を思案中
- ❖ =>対応は富山の国際会議以降になる