



Analytics Design Lab

Nomolytics

特許文書データの分析事例

株式会社アナリティクスデザインラボ

1. Nomolyticsについて	3
2. Nomolyticsを適用した特許分析事例	17
3. Nomolyticsを適用した特許分析のメリット	44

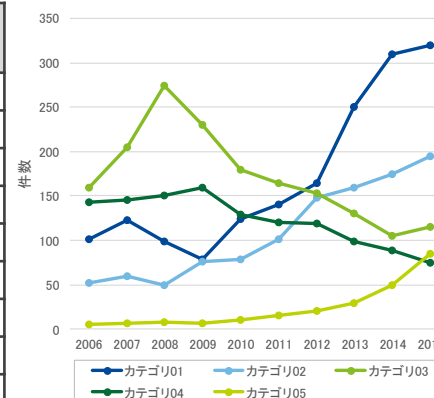
1. Nomolyticsについて

1-1. 従来の特許文書分析の課題とAI技術の応用

単語をベースに、あるいは手動でグルーピングしたカテゴリをベースに、全体の出現状況、経年変化、出願人の特徴、課題と解決手段の関係などを把握する分析がよく行われます

- ## 手動設定したカテゴリのトレンド把握

例) 掃除機カテゴリーのリスト
掃除機
集塵
集塵容器
吸引力
サイクロン
塵埃->分離
塵埃->吸い込む
塵埃->収容
塵埃->遠心分離



- 抽出した単語を手動でいくつかのカテゴリにグルーピングする
- 各カテゴリの出現年ごとの出現頻度をグラフ化し、トレンドを把握する

- 単語の出現データから共通して現れる特徴的な軸を2つ抽出する
- その2軸による平面上に単語と出願人を同時にマッピングする
- 出願人の周辺に配置された単語群から各出願人の特徴を考察する

課題	解決手段カテゴリ						
	カテゴリ01	カテゴリ02	カテゴリ03	カテゴリ04	カテゴリ05	カテゴリ06	カテゴリ07
カテゴリ01	206	80	71	184	26	47	11
カテゴリ02	208	76	87	182	23	48	9
カテゴリ03	172	74	53	57	31	35	10
カテゴリ04	176	54	37	59	26	46	29
カテゴリ05	85	39	13	23	14	16	5
カテゴリ06	87	53	31	33	59	37	15

- 「要約」の【課題】と【解決手段】それぞれに対して出現単語のカテゴリを設定する
- 課題と解決手段のカテゴリのクロス集計をして、用途と技術の関連性を考察する

これまでの特許文書分析の課題と解決アプローチ

2つのAI技術を組み合わせることで、特許文書データを単語ベースではなく、客観的に抽出されるトピックベースで解釈し、そのトピックの統計的な関連性を分析できます

課題①

単語ベースの分析では
複雑で考察しにくい

課題②

カテゴリの設定が主観的で
作業負荷も大きい

課題③

課題と解決手段の統計的な
関係を分析していない

単語を賢くクラスタリングする
AI技術

PLSA

確率的潜在意味解析

使われ方の似ている単語群を
トピックとして集約する

要因関係をモデリングする
AI技術

ベイジアンネットワーク

抽出したトピックに関わる要因
関係を統計的にモデル化する

PLSAは、トピックモデルと呼ばれるAI技術で、テキストマイニングで抽出された大量の単語をいくつかのトピックに集約して類型化することができます

PLSAの概要

- 行列データの行の要素xと列の要素yの背後にある共通特徴となる潜在クラスzを抽出する手法である
- 元々は文書分類のための手法として開発されている (Hofman, 1999)
- 各文書の出現単語を記録した文書(行) × 単語(列) という高次元(列数の多い)共起行列データに適用して複数の潜在トピックを抽出し、文書(行) × トピック(列) という低次元データに変換して文書を分類する

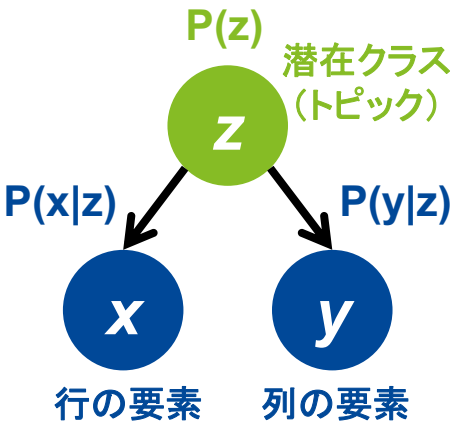
「文書×単語」行列（共起行列）

文書ID	単語1	単語2	単語3	...	単語5,014	単語5,015
1	0	0	1		1	0
2	1	0	2		0	1
...						

文書ID	トピック1	トピック2	...	トピック15
1	0.09%	0.03%		0.04%
2	0.01%	0.12%		0.06%
...				

例えば数千列ある高次元のデータでも十数個の潜在トピックで説明することができる

PLSAのグラフィカルモデル



- $P(x|z)$, $P(y|z)$, $P(z)$ の3つの確率が計算される
- 潜在クラスzの数はあらかじめ設定する

※条件付確率 $P(A|B)$
事象Bが起こる条件下で事象Aの起こる確率

xとyの共起確率を潜在クラスzを使って表現する

$$P(x, y) = \sum_z P(x|z)P(y|z)P(z)$$

PLSAのメリット

行の要素と列の要素を同時にクラスタリングできる

潜在クラスは行の要素と列の要素の2つの軸の変動量に基づいて抽出され、結果も2つの軸の情報から潜在クラスの意味を解釈することができる

ソフトクラスタリングできる

全ての変数が全てのクラスに所属し、その各所属度合いが確率で計算されるため、複数の意味を持つ変数がある場合でも自然と表現できる

複雑な観測情報をシンプルにかつ忠実に把握するため、PLSAを選択します

階層型 クラスタ分析

- Ward法など
- 要素間の距離を計算し、距離の近い要素同士を結合してクラスタを構成していく
- 結合の過程が樹形図で表され、結果を見てからクラスタ数を決められる(ボトムアップ的なクラスタ分析)
- データ数が多くなると計算が膨大となる

非階層型 クラスタ分析

- k-means法など
- あらかじめクラスタ数を決め、そのクラスタ数に全要素を一回でグループリングする
- 各クラスタ(の重心)に対して要素の距離を計算し、距離の近い要素で集められたクラスタとなるように分類結果を調整する
- 階層型クラスタ分析よりも計算量が抑えられる

LSA (Latent Semantic Analysis)

- 特異値分解と呼ばれる
- $(m \times n)$ の行列を、 $(m \times k), (k \times k), (k \times n)$ に分解する
- m 個のデータと n 個の変数を、 k 個の潜在クラスで表現する(クラス数はあらかじめ設定)
- 大きな値をとりやすいクラスが残る傾向にあるため、各要素は事前に重み付けする必要がある

PLSA (Probabilistic Latent Semantic Analysis)

- LSAを確率的に処理
- LSAのような事前の重み付けは必要がない
- $P(x, y)$ の確率を、 $P(x|z), P(y|z), P(z)$ に分解する
- 行要素 x と列要素 y を、潜在クラス z で表現する(クラス数はあらかじめ設定)
- 結果は観測データのみから定義され、新規データはクラスで表現できない(過学習)

LDA (Latent Dirichlet Allocation)

- PLSAをベイズ拡張した手法
- PLSAの過学習の問題に対して、LDAはディレクレ分布を事前分布に仮定し新規データのクラスを推定できる
- 新規データに対応するため、抽出されるクラスは観測データを忠実に再現するものではなく、クラスの抽象度が高い傾向がある

従来のクラスタ分析

- 基本的に要素間の距離に基づいて分類を行う
- 要素数が多くなると要素間の距離が離れていき妥当な結果が得られにくい(次元の呪い)
- 列要素の距離に基づいて行要素を分類するか、行要素の距離に基づいて列要素を分類し、行と列どちらか一方を分類する
- 一つの要素は必ず一つのクラスタに所属し、重複所属を許さないハードクラスタリングとなる

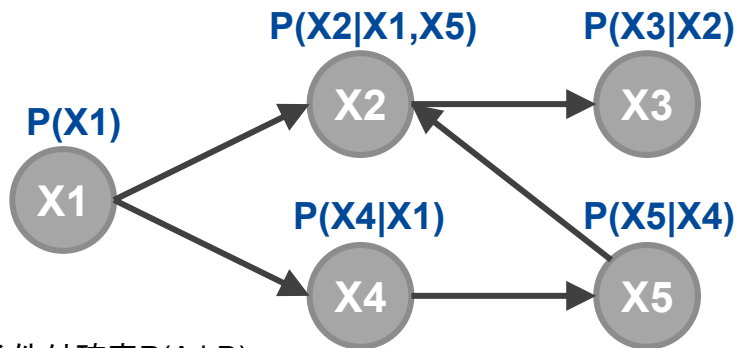
トピックモデル

- 単語一つ一つが列の要素となる超高次元のテキストデータを想定した手法
- 要素間の距離の近さで分類するのではなく、高次元データの情報をできるだけ保存した形で低次元に変換する次元圧縮手法であるため、要素数が多い複雑なデータにも対応できる
- 行の要素と列の要素の背後にある共通する特徴をクラスとして抽出するため、行と列の両方をクラスタリングでき、クラスの持つ情報が多い
- 一つの要素は全てのクラスに所属するソフトクラスタリングで、その所属の重みを計算するため、データが複数の特徴をまたがる場合でも表現できる

ベイジアンネットワークは、ベイズ推論に基づいたAI技術で、変数と変数の間に潜む確率的な因果関係をネットワーク構造で探索することができます

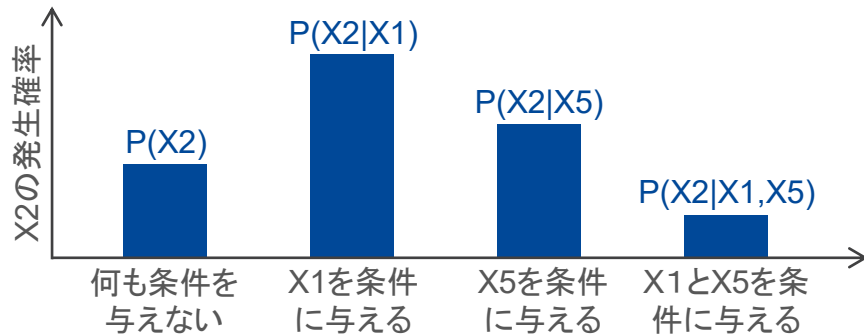
ベイジアンネットワークの概要

- 複数の変数の確率的な因果関係をネットワーク構造で表わし、ある変数の状態を条件として与えたときの他の変数の条件付確率を推論することができる
- 目的変数と説明変数の区別はなく、様々な方向から変数の確率シミュレーションができる
- 全ての変数は質的変数(カテゴリカル変数)となるため、量的変数の場合は閾値を設けてカテゴリに分割する
- 確率論の非線形処理によるモデル化のため、非線形の関係や交互作用が生じる現象でも記述できる



※条件付確率 $P(A|B)$
事象Bが起こる条件の下で事象Aの起こる確率

確率的因果関係と交互作用



- X2の発生確率は、何も条件を与えない時(事前確率)と比べて、X1やX5を条件に与えると確率が上昇する
⇒X1やX5はX2の発生に関して“確率的な”因果関係がある
- しかし、X1とX5の両方を条件に与えると、元々の事前確率よりも確率が下がってしまう
⇒X1とX5はX2に対して交互作用がある(X1とX5は相性が悪い)

ベイジアンネットワークのメリット

現象を理解して柔軟にシミュレーションできる

目的変数、説明変数の区別なく変数の関係をモデル化するので、現象の構造を理解でき、推論変数と条件変数を自由に指定して確率推論できる

効果を発揮する有用な条件を発見できる

ある条件のときにだけ効果が現れるといった交互作用がある場合でも、確率的に意味のある関係としてモデル化することができる

テキスト情報内に潜む要因関係を理解するため、ベイジアンネットワークを選択します

ニューラルネットワーク (ディープラーニング)

- 入力(説明変数)と出力(目的変数)の関係(非線形)をモデル化する
- 入力と出力の間に中間層(隠れ層)を設定し、入力情報に重みをつけて出力精度を高める処理を中間層で行う
- 柔軟性が高く複雑な関係もモデル化でき予測精度も高まるが、処理が複雑すぎてモデルの中身がブラックボックス化してしまう

回帰分析・判別分析 (数量化 I 類・II 類)

- 目的変数を説明変数の1次結合で定式化する
- 目的変数と説明変数の間に線形関係があるという仮定に基づいている
- 各説明変数の影響は独立しており、複合的な交互作用の影響は表現できない
- 説明変数間で相関が高い場合は解が不安定となり(多重共線性)、変数が多い場合この解消検討の負荷が大きい

決定木

- 目的変数の特徴がよく現れる条件ルールを説明変数とその閾値による分岐で構成する
- ルールがツリー構造で可視化されるため目的変数と各説明変数の関係が分かりやすい
- 目的変数と説明変数の非線形な関係もモデル化でき、複合条件で効果が変化する交互作用を表現しやすい

ベイジアンネットワーク

- 複数の変数の確率的な因果関係をネットワーク構造でモデル化する
- 目的変数と説明変数の区別がないため、それぞれの変数が互いにどのような関係をもってそのデータの現象を構成しているのか理解できる
- 変数間の関係は条件付確率で計算され、複合条件によって効果が変化する交互作用も表現できる

モデルの構造が不明

モデルの構造(要因関係)が理解できる

非線形のモデル化

線形のモデル化

非線形のモデル化

目的変数と説明変数の区別がある

区別がない

AIも様々な技術があり、例えば「理解系AI」「識別系AI」「生成系AI」に分類することができますが、それぞれの技術を分析目的に応じて賢く使いこなすことが求められます

理解系AI

現状のデータに潜む特徴や要因関係を理解するAIであり、ホワイトボックスのモデルが求められる

PLSA

LDA

決定木

ベイジアンネットワーク

「理解系AI」により、データに潜む傾向を人間が理解し、ビジネスアクションを人間が考え実行する

※私見による分類であり、一般的に定義された分類ではありません

識別系AI

画像判定や文章分類など、新規のデータを識別するAIであり、精度さえ良ければモデルはブラックボックスでもよい

深層学習・CNN

RNN・LSTM・NMT

Transformer・T5

BERT

生成系AI

入力した情報に対して画像や文章を生成するAIであり、精度さえ良ければモデルはブラックボックスでもよい

「識別系AI」や「生成系AI」は人間がデータを理解して業務に活用するのではなく、業務の自動化・省力化を目的としていることが多い

GAN

GPT

Diffusion model

1. Nomolyticsについて

1-2. Nomolyticsの概要と適用対象

膨大なテキストデータをトピックに変換して解釈を容易にし、テキスト情報内に潜む要因関係をモデル化して、ビジネスアクションに有用な特徴を把握可能にします

Nomolytics : Narrative Orchestration Modeling Analytics

テキストマイニング

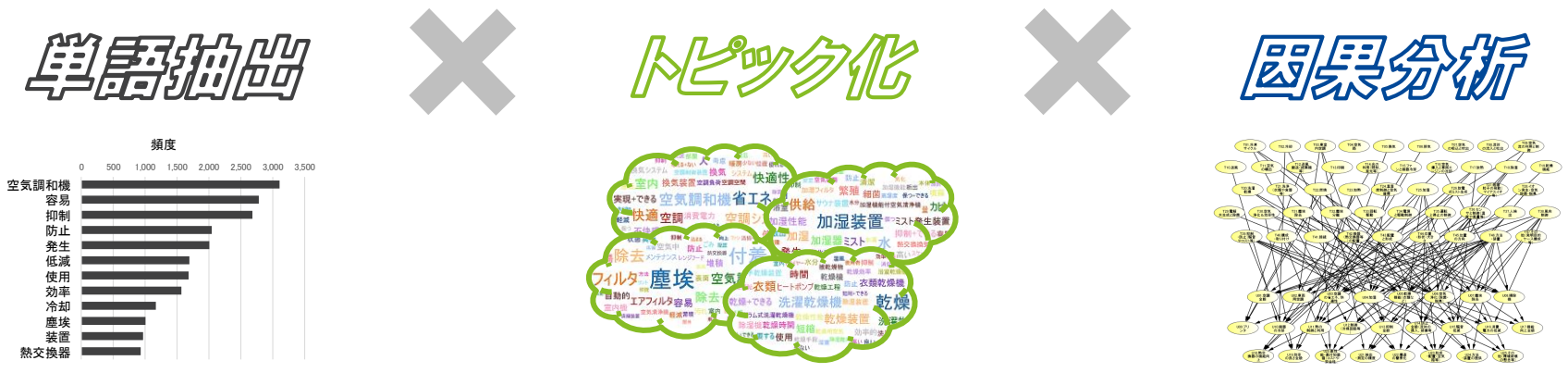
文章に含まれる単語を抽出し、その出現頻度を集計する

PLSA
確率的潜在意味解析

単語が出現する特徴を学習し、膨大な単語を複数のトピックにまとめる

ベイジアンネットワーク

トピックやその他属性情報など、テキスト情報内の要因関係をモデル化する



Nomolyticsのメリット

- 膨大なテキストデータをいくつかのトピックという人間が理解しやすい形に整理し類型化できる
- テキスト情報に潜む要因関係を構造化し、特徴を見たいターゲットのキーワードを発見できる
- 条件を変化させたときの効果を確率的にシミュレーションでき、有効なアクションを検討できる

Nomolyticsは様々な業務のテキストデータに適用することができます



口コミ

- 顧客ターゲット別の関心トピックを把握
- 製品・サービス別のトピックを把握
- 口コミ得点に寄与するトピックを把握
- ニーズに応じたマーケティングを検討



アンケート

- 自由記述回答の内容をトピックで把握
- トピック化された自由記述回答と通常の定型設問回答の関係を統計分析
- 顧客満足度に寄与するトピックを把握



コールセンター履歴

- 問い合わせ内容をトピックで把握
- 製品別・顧客別のトピック傾向を把握
- 解約・退会に寄与するトピックを把握
- 満足度向上、顧客離反抑制の施策検討



特許文書

- 特許文書の内容をトピックで把握
- ティンドや競合他社の動向を把握
- 用途と技術の関係分析から用途実現の技術戦略や保有技術の新規用途を検討



営業日報

- 営業活動内容をトピックで把握
- 営業属性別のトピック傾向を把握
- 成約に寄与するトピックを把握
- 成約のための効果的な営業教育を検討



有価証券報告書

- 企業・業界の事業内容をトピックで把握
- 事業内容トピックのトレンドを把握
- 好業績に寄与する事業トピックを把握
- 定性情報から行う企業分析・業界分析



エントリーシート

- 志望動機やPR文のトピックを把握
- 記述トピックに基づいて学生を分類
- トピック傾向から面接の質問内容を検討
- 選考通過に寄与するトピックを把握



診療・看護記録

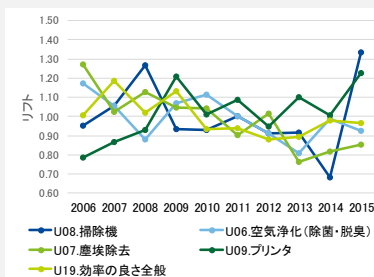
- 診療記録、看護記録をトピックで把握
- 患者の属性別のトピック傾向を把握
- 検査指標に寄与する定性情報を把握
- 定性情報も用いた診療・助言を検討



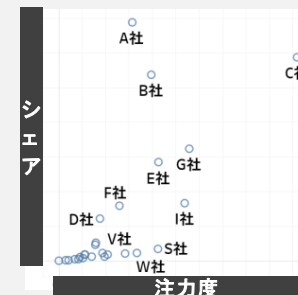
問題発生レポート

- 不具合やヒヤリハットをトピックで整理
- 作業環境別のトピック傾向を把握
- 重大問題に寄与するトピックを把握
- 問題を抑制する作業・環境改善を検討

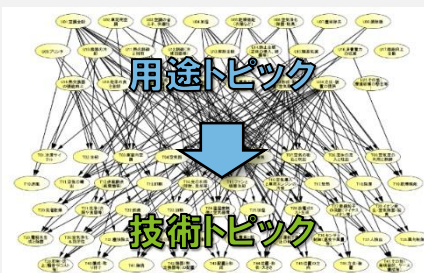
出願年×トピックの関係を分析することで、用途や技術のトレンドを把握します



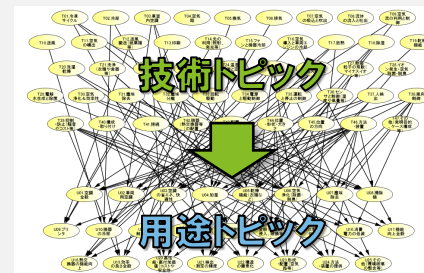
出願人×トピックの関係を分析することで、各社の特徴やポジションを把握します



用途に対する技術の関係を分析することで、ある用途を実現する上で重要な技術や各社の出願動向を把握し、自社の開発戦略や他社との協業可能性を探ります



技術に対する用途の関係を分析することで、自社技術と関係がある用途のうち想定をしていない用途を発見し、技術の新規用途展開のアイデアを創出します



斤

2. Nomolyticsを適用した特許分析事例

2-1. 「風・空気」に関する特許文書データ

「風」「空気」に関する10年分の特許データ30,039件の要約に記載されている【課題】と【解決手段】の文章を分析します

データの抽出条件と抽出結果

- 対象
 - 公開特許公報
- キーワード
 - 要約と請求項に「風」と「空気」を含む
- 出願年
 - 2006年1月1日～2015年12月31日
- 抽出方法
 - PatentSQUAREを使用
- 抽出結果
 - 30,039件



分析データの加工

- 要約文の【課題】と【解決手段】に記載されている文章をそれぞれ抽出する
 - このような書式で記載されていないものは要約文をそのまま使用する
- 出願人情報は名寄せをし、グループ会社などは統一する

課題の文章

【要約】【課題】ユーザーの快適性を維持しつつ、省エネ運転を行うことができる空気調和機を提供すること。【解決手段】本発明の空気調和機は、室内温度を検出する室内温度検出手段と、人体の活動量を検出する人体検出手段と、基準室内設定温度を設定するリモコン装置30とを備え、室内温度が基準室内設定温度となるように空調制御を行う空気調和機であって、人体検出手段で検出する活動量が所定の活動量以内であるときは、室内温度が、基準室内設定温度を補正した補正室内設定温度となるように空調を行い、補正室内設定温度よりも低い状態を継続すると、圧縮機を停止させ、圧縮機の復帰は、基準室内設定温度に基づいて行う。

解決手段の文章

2. Nomolyticsを適用した特許分析事例

2-2. Nomolyticsを適用した特許分析のプロセス

特許要約の【課題】と【解決手段】の文章から用途と技術のトピックを抽出し、各トピックのトレンドや出願人の特徴分析、また用途と技術の関係分析によって技術戦略を検討します

トピックの抽出

特許の要約文に**テキストマイニング**を実行して単語や係り受け表現を抽出

単語	品詞	頻度
空気調和機	名詞	3,106
空気	名詞	2,846
容易な	形容詞	2,790
抑制する	動詞	2,687
塵埃	名詞	1,687
分離する	動詞	1,231
...	...	...

PLSAの適用

【課題】の要約文から用途トピックを抽出

掃除機 加湿 空気清浄 除菌

【解決手段】の要約文から技術トピックを抽出

塵埃分離 除湿 イオン発生

トピックのスコアリング

全特許データに対する各トピックの該当度を計算

ID	出願年	出願人	用途トピック1	用途トピック2	用途トピック*	技術トピック1	技術トピック2	技術トピック*
1	2014	A社	2.1	0.6	...	1.5	5.0	...
2	2013	B社	0.3	3.4	...	4.6	0.9	...
3	2011	C社	4.8	2.2	...	2.7	1.1	...
n	...	...	...	...	...	...	...	...

トレンドの分析

トピックのスコアを出願年で集計することでトピックのトレンドを把握

競合他社の分析

トピックのスコアを出願人で集計することで各社の特徴やポジションを把握

用途と技術の関係分析

用途トピックと技術トピックの関係性を**ベイジアンネットワーク**でモデル化

技術トピック

用途トピック

用途⇒技術 の分析

検討中の用途を実現する上で重要な技術と各社の動向を把握し、**事業化のための技術戦略**を検討する

技術⇒用途 の分析

自社技術と関係する用途のうち未想定用途を発見し、**技術の新規用途展開**のアイデアを創出する

2. Nomolyticsを適用した特許分析事例

2-3. トピックの抽出

トピック抽出のアプローチ

テキストマイニングで単語と係り受け表現を抽出し、単語 × 係り受けで構成される共起行列にPLSAを適用することで単語と係り受けの出現の背後にある潜在トピックを抽出します

テキストマイニングの実行

【課題】と【解決手段】の文章に含まれる単語と係り受けを抽出する

単語	品詞	頻度
空気調和機	名詞	3,106
空気	名詞	2,846
容易	名詞	2,790
抑制	名詞	2,687
良い	形容詞	2,481
向上	名詞	2,328
防止	名詞	2,047
発生	名詞	2,005
...	...	...

係り受け表現	頻度
空気調和機⇒提供	1,575
効率⇒良い	1,325
車両用空調装置⇒提供	578
掃除機-提供	545
容易-構成	539
画像形成装置-提供	334
抑制-提供	296
向上-図る	279
...	...

共起行列の作成

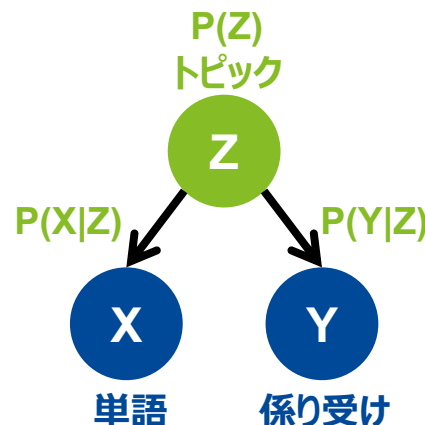
抽出した単語と係り受け表現に基づいて、「単語 × 係り受け」の共起行列(文章単位で同時に出現する頻度のクロス集計表)を作成する

		係り受け表現				
単語		空気調和機↓提供	効率↓良い	車両用空調装置↓提供	掃除機↓提供	...
	空気調和機	1578	100	4	1	
	空気	85	144	45	50	
	容易	100	105	51	67	
	抑制	142	95	64	63	
	...					

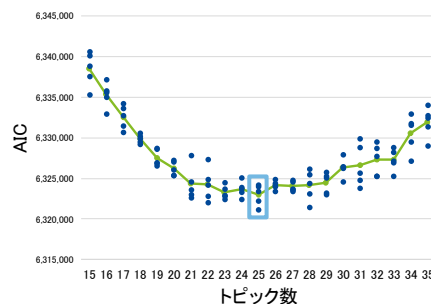
共起行列の構成(それぞれ頻度10件以上を対象)
課題: 単語(3,256語) × 係り受け(2,084表現)
解決手段: 単語(5,187語) × 係り受け(7,174表現)

PLSAの実行

共起行列にPLSAを適用する



トピック数を幅を持たせて設定し、各トピック数に対してPLSAを初期値を変えて5回ずつ実行して情報量基準AICを計算し、AIC最小の解を採用する



トピックの抽出

各トピックについて以下の3つの確率が計算される

- ① $P(X|Z)$
トピックにおける単語の所属確率
- ② $P(Y|Z)$
トピックにおける係り受けの所属確率
- ③ $P(Z)$
トピックの存在確率

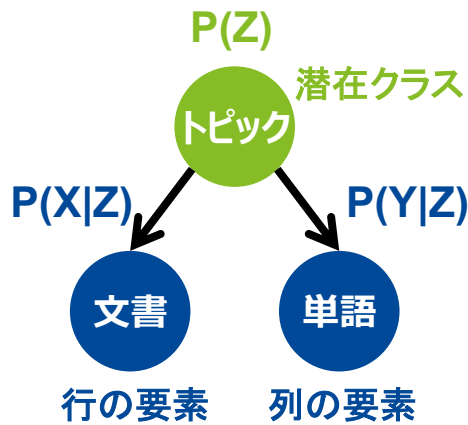
トピックにおける $P(X|Z)$ と $P(Y|Z)$ からトピックの意味を解釈する

トピック T32			
$P(Z) = 2.7\%$			
$P(X Z)$	単語	$P(Y Z)$	係り受け
5.5%	送風機	2.1%	塵埃-分離
5.2%	塵埃	1.7%	分離-塵埃
4.1%	掃除機	1.7%	塵埃-含む
3.6%	分離	1.5%	吸い込む-塵埃
3.5%	吸い込む	1.3%	含む-空気
2.3%	集塵部	1.0%	空気-分離
1.9%	配置	1.0%	送風機-吸い込む
1.9%	集塵容器	1.0%	発生-送風機
1.6%	旋回	0.9%	含塵空気-分離
1.5%	含塵空気	0.9%	備える-掃除機
...	...	...	...

確率の高い構成要素から、トピックT32は「塵埃の分離」に関するトピックと解釈できる

PLSAのインプットとする共起行列の構成を「文書 × 単語」ではなく「単語 × 係り受け」とすることで、要素間の違いが出やすくなり、解釈のしやすいトピックを抽出できます

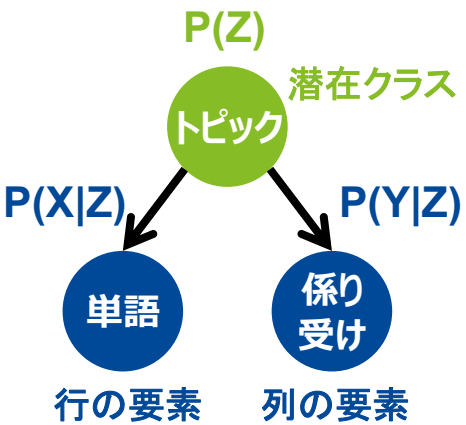
一般的なPLSAの共起行列



	単語1	単語2	単語3	単語4	...
文書ID:1	1	1	0	0	
文書ID:2	0	0	2	0	
文書ID:3	0	0	0	1	
文書ID:4	2	0	0	0	
...					

- 共起行列はBag-of-Wordsによる単語の頻度で構成され、ほとんどが“0”となる疎なデータであるため、要素間の違いが現れにくく、クリアなトピックを抽出しにくい
- PLSAのトピックには行の要素と列の要素が同時に所属し、両方の情報軸からトピックの意味を解釈できるが、一方の軸(行)は文書IDという意味性の低い情報で、トピックの解釈に使用しにくい

NomolyticsでのPLSAの共起行列

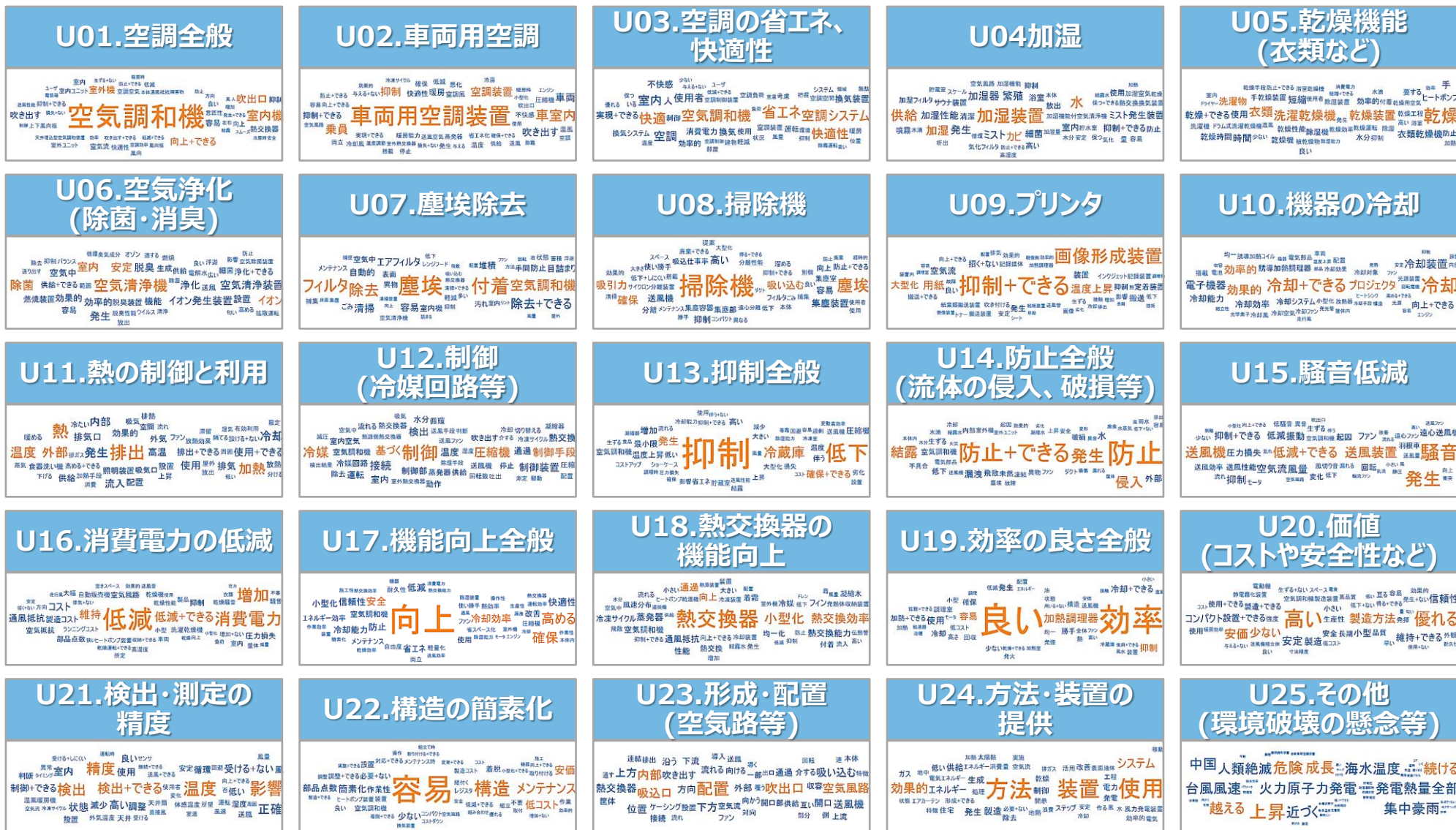


	係り受けa	係り受けb	係り受けc	係り受けd	...
単語1	325	264	11	20	
単語2	241	201	6	8	
単語3	28	41	288	14	
単語4	9	15	4	172	
...					

- 共起行列はクロス集計型の行列で、単語と係り受けの共起頻度が入った密なデータであるため、要素間での違いが現れやすく、クリアなトピックを抽出しやすい
- 行と列が単語と係り受けで構成されている共起行列では、どちらもそれぞれ単独で意味を持つ情報となるため、両方の情報軸からトピックの意味を解釈することができ、解釈の容易性が高まる

用途トピック25個の一覧

【課題】の文章からは、空調や加湿、空気清浄、掃除機、プリンタ、機器冷却、騒音や消費電力の低減、構造の簡素化などの用途が25個抽出されました



※文字の大きさはトピックに対する関係の強さを表現している
(上位5つの単語を赤色で表示している)

【解決手段】の文章からは、空気の冷却や空気路、換気、放熱、除湿、乾燥、加湿、イオン生成、空気清浄、塵埃分離、センサと制御、構成や配置などの技術が47個抽出されました



※文字の大きさはトピックに対する関係の強さを表現している
(上位5つの単語を赤色で表示している)

24 Nomolytics

2. Nomolyticsを適用した特許分析事例

2-4. トピックのスコアリング

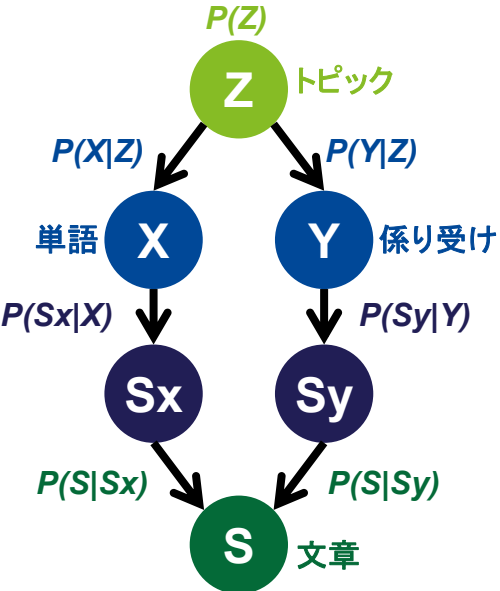
トピックのスコアリング

文章単位に各トピックのスコア(該当度)を計算し、それを特許ID単位に集約し、最終的には閾値を設定して{1:該当有, 0:該当無}のデータに変換します

文章単位
のスコア

$$\frac{P(S|Z)}{P(S)}$$

- リフト値(事後確率÷事前確率)
- トピックを条件とすることで文章の発生確率が何倍になるのかを示す



文章を単語で定義される文章 S_x と係り受けで定義される文章 S_y を設定し、それぞれトピックとの関係を計算し、最終的にそれらを一つに統合する

単語 X_i で定義される文章 Sx_h
$Sx_h = \{X_1, X_2, \dots, X_i\}$
トピック Z_k を条件とした文章 Sx_h の出現確率
$P(Sx_h Z_k) = \sum_i P(Sx_h X_i)P(X_i Z_k)$
単語 X_i が出現する中で文章 Sx_h が出現する確率(X_i の出現文章数の逆数)
$P(Sx_h X_i) = 1/n(X_i)$
係り受け Y_j で定義される文章 Sy_h
$Sy_h = \{Y_1, Y_2, \dots, Y_j\}$
トピック Z_k を条件とした文章 Sy_h の出現確率
$P(Sy_h Z_k) = \sum_j P(Sy_h Y_j)P(Y_j Z_k)$
係り受け Y_j が出現する中で文章 Sy_h が出現する確率(Y_j の出現文章数の逆数)
$P(Sy_h Y_j) = 1/n(Y_j)$
トピック Z_k を条件とした文章 S_h の出現確率 ※ $P(S_h Sx_h)$ と $P(S_h Sy_h)$ はともに1/2とする
$P(S_h Z_k) = P(S_h Sx_h)P(Sx_h Z_k) + P(S_h Sy_h)P(Sy_h Z_k)$
文章 S_h の出現確率
$P(S_h) = \sum_k P(S_h Z_k)P(Z_k)$

トピックスコア算出プロセス

①文章ごとにスコアを計算

特許ID	文章ID	T01	T02	T03	...	T47
1	1	3.1	0.9	2.0		1.1
1	2	1.4	0.2	5.5		2.4
2	1	0.8	5.8	1.3		0.9
2	2	1.2	3.2	1.7		1.0
2	3	0.6	1.8	2.6		3.6
...						

②特許IDごとに文章スコアを集約

※最大値を採用する

特許ID	T01	T02	T03	...	T47
1	3.1	0.9	5.5		2.4
2	1.2	5.8	2.6		3.6
...					

③閾値を設定してフラグに変換する

※閾値は3に設定する

特許ID	T01	T02	T03	...	T47
1	1	0	1		0
2	0	1	0		1
...					

トピックのフラグデータの作成

全特許データに対して各トピックのスコア(該当有無のフラグ情報)を計算することで、トピックをベースとした様々な分析を実行することができます

トピックのスコア(フラグ情報)を紐づけた特許データ

特許ID	出願番号	出願年	出願人	要約文		用途トピック U01	用途トピック U02	...	用途トピック U25	技術トピック T01	技術トピック T02	...	技術トピック T47
				【課題】	【解決手段】								
1	特願2006-XXXX	2006	A社	空気調和機の高外気	吸気口から導入された	1	0		0	0	1		0
2	特願2009-XXXX	2009	B社	短時間で除霜を行うこ	着霜検出手段が室外	0	1		0	1	0		0
3	特願2011-XXXX	2011	C社	乾燥運転が中断され	通風路を通して回転槽	0	0		1	1	0		0
4	特願2013-XXXX	2013	D社	ウインドシールドの防	車両用空調装置の空	0	1		0	0	1		1
...	...	...	...			...	...		...	...	...		...
30039	特願2012-XXXX	2012	Z社	プリ空調時に、除菌ま	冷暖房空調ユニットは	0	1		0	1	1		0

①出願年の分析

②出願人の分析

③用途と技術の関連性の分析

トピックをベースにした分析によって読むべき
特許文書を効率的に絞り込むことができる

2. Nomolyticsを適用した特許分析事例

2-5. 出願年×トピックによるトレンド分析

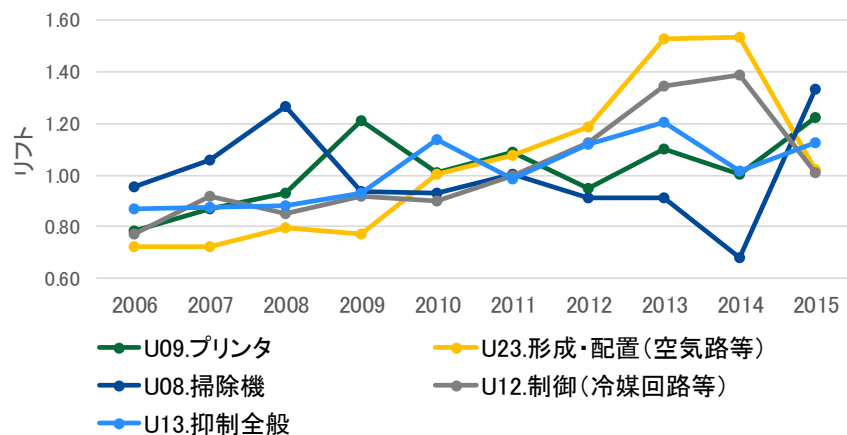
【分析目的】

技術や用途のトレンドを把握し、有望なシーズやニーズを探り、今後の技術開発戦略を検討する

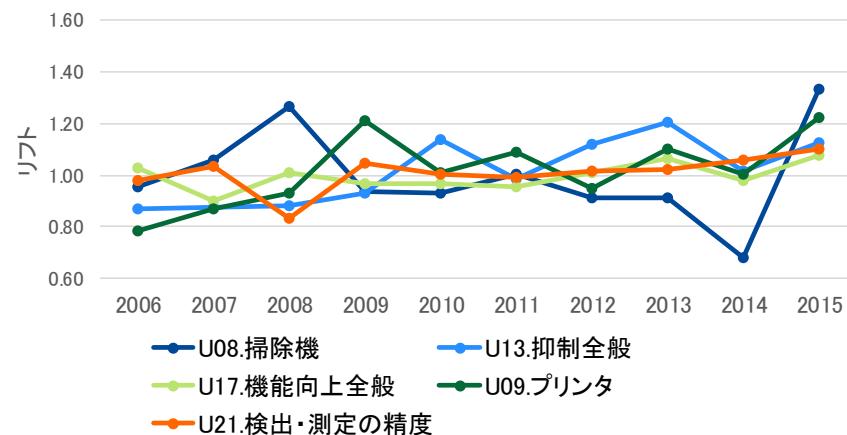
用途トピックの上昇トレンド

近年は掃除機や空気浄化、塵埃除去、プリンタに関する用途が上昇しています

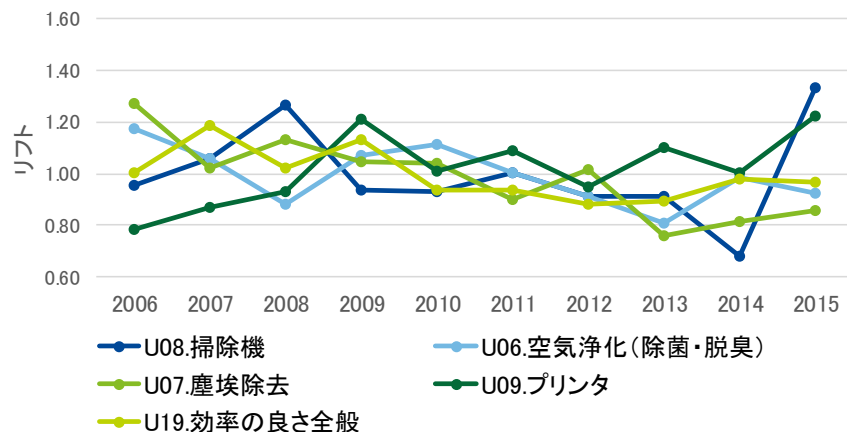
【長期】 2006年からの上昇率 best5



【中期】 2011年からの上昇率 best5



【短期】 2013年からの上昇率 best5



集計の仕方

■ リフト値を出願年・トピックごとに集計

$$P(\text{出願年} \mid \text{トピック } T_x = 1)$$

$$P(\text{出願年})$$

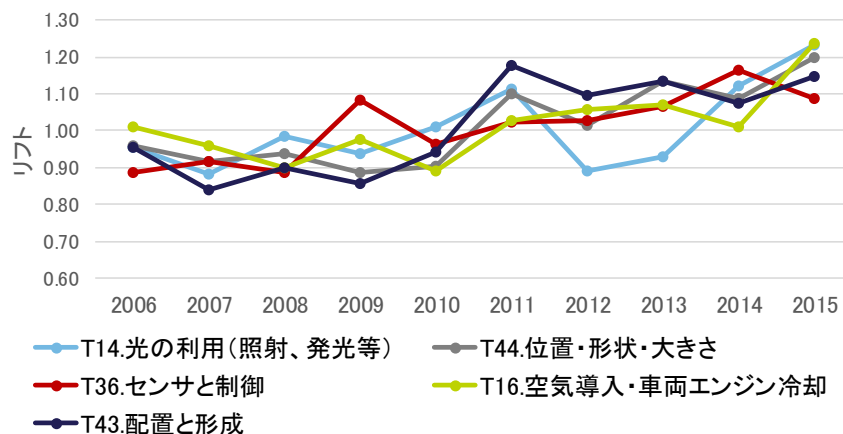
■ リフト値は出願年とトピックの関係を示す指標

■ トピック毎の各出願年の出願割合を、その出願年の出願割合で正規化した値

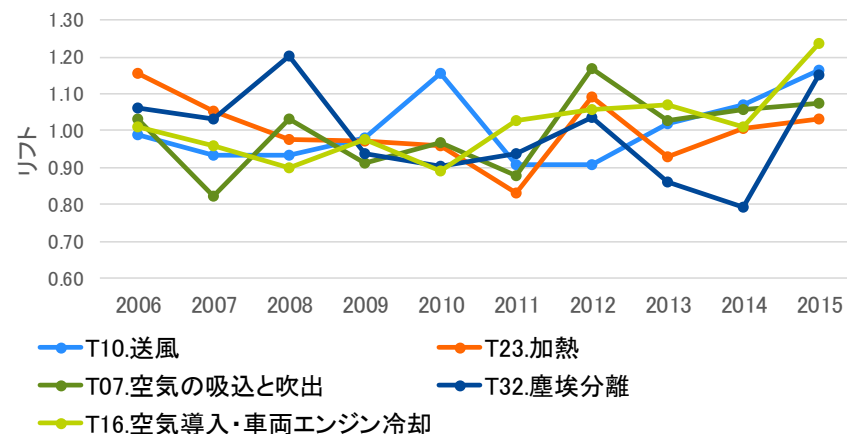
技術トピックの上昇トレンド

近年は塵埃分離や車両エンジンの冷却に関する技術が、長期的にはプロジェクタなどの光の利用に関する技術が上昇しています

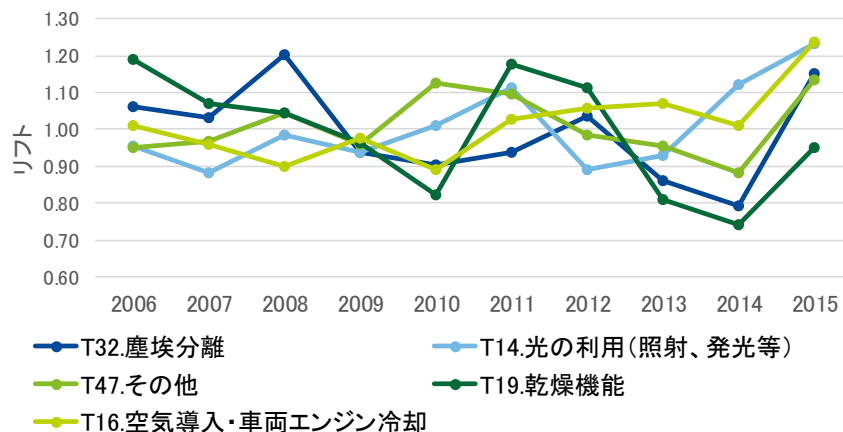
【長期】 2006年からの上昇率 best5



【中期】 2011年からの上昇率 best5



【短期】 2013年からの上昇率 best5



集計の仕方

- リフト値を出願年・トピックごとに集計

$$P(\text{出願年} \mid \text{トピック} T_x = 1)$$

$$P(\text{出願年})$$

- リフト値は出願年とトピックの関係を示す指標

- トピック毎の各出願年の出願割合を、その出願年の出願割合で正規化した値

2. Nomolyticsを適用した特許分析事例

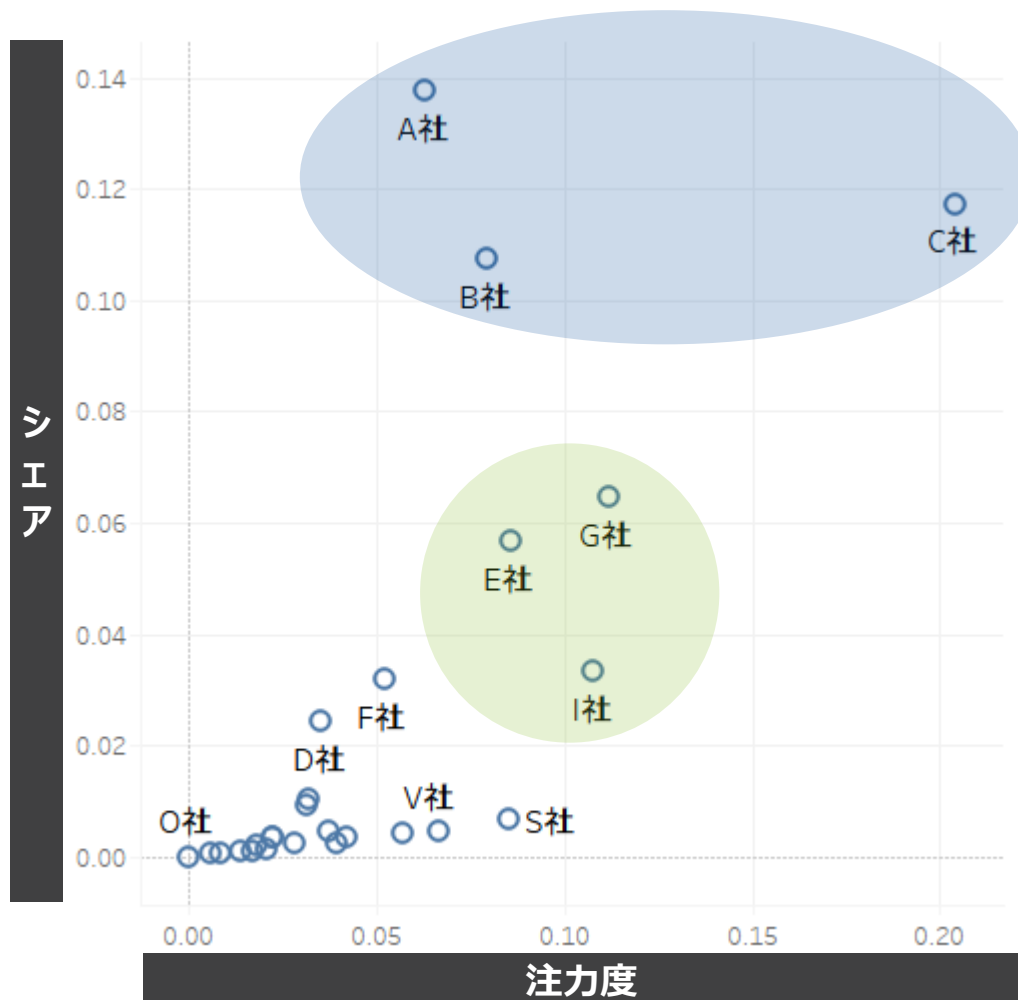
2-6. 出願人×トピックによる競合分析

【分析目的】

各出願人の動向や、業界における棲み分け、競合他社と自社との関係性などを把握し、他社との差別化戦略や協業戦略を検討する

塵埃分離に関する技術は、3社のシェアが高いものの、他にもある程度のシェア・注力度を有する企業が何社か存在するため、連携によって競争力を高める動きも考えられます

注力度とシェアの散布図



考察と戦略の検討

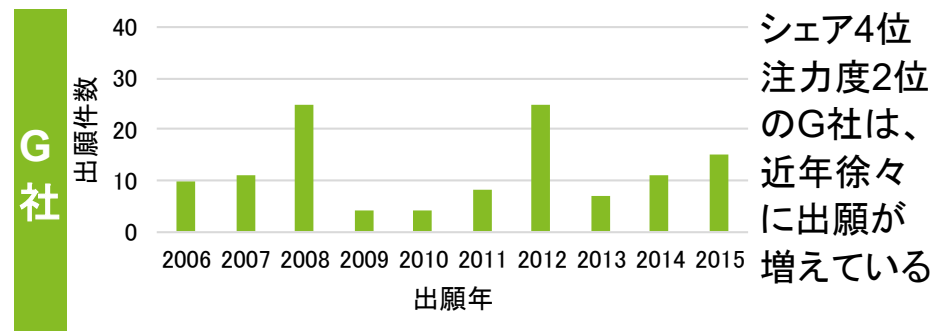
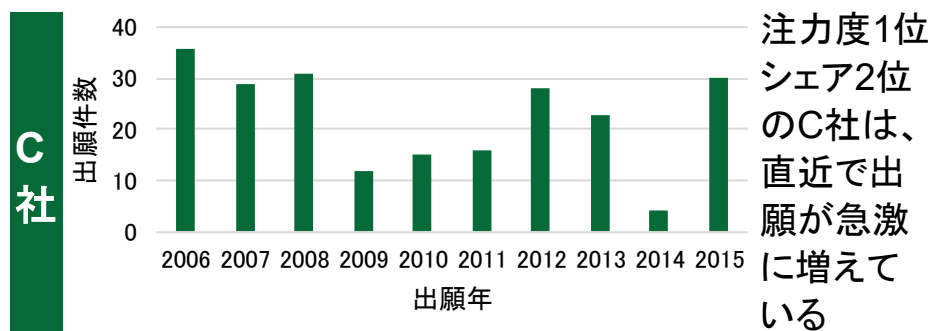
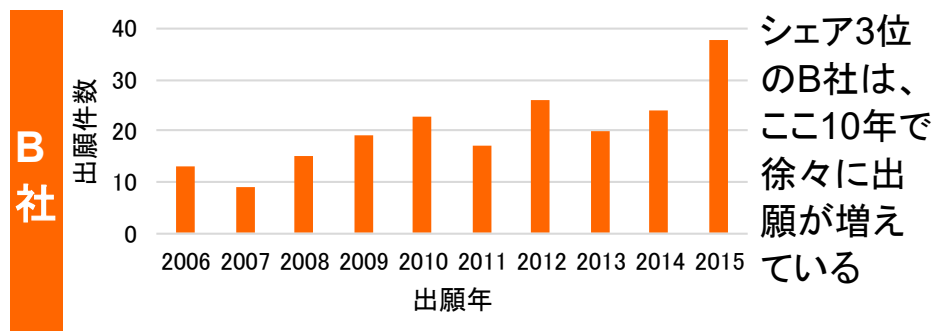
- シェアではA社・B社・C社が高いが、特にC社は注力度がとても高く、特有の技術力を保有していると考えられる
- E社・G社・I社はシェアは中程度だが、注力度は比較的高く、技術力もあると思われる
- 高いシェアを持つ企業は、中程度のシェアの企業と連携することで、より技術力を高めながらシェアを伸ばすことが期待できる
- あるいは中程度シェアの企業の間で連携し、高シェアの業界大手に対抗することも考えられる

注力度とシェア

- **注力度**: $P(\text{トピック}T \mid \text{出願人}X)$
 - 出願人Xの出願特許の中で、どれくらいの割合がそのトピックTに該当するものか、つまり出願人がどれくらいそのトピックに注力しているのかを示す
- **シェア**: $P(\text{出願人}X \mid \text{トピック}T)$
 - トピックTが該当する特許の中で、どれくらいの割合がその出願人Xの出願によるものか、つまりトピックの中でどれくらいその出願人が占めているのかを示す

高シェアのA社とB社の近年の出願動向は、A社は減少ですがB社は増加し、注力度1位のC社は直近で出願が急増し、シェア4位のG社も出願を伸ばしており、今後に要注目です

注目企業の出願件数の推移



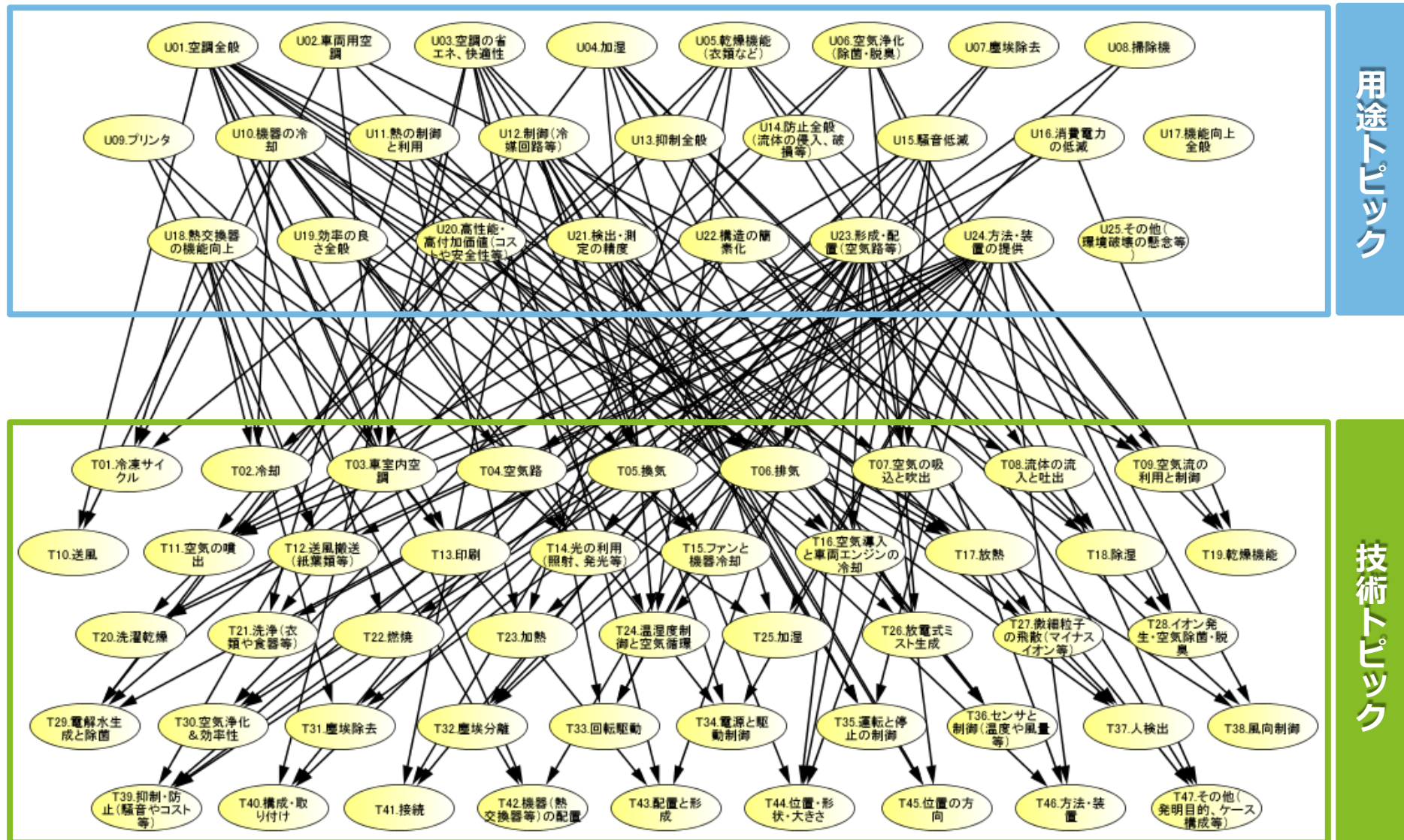
2. Nomolyticsを適用した特許分析事例

2-7. 用途×技術の関係分析＜その1＞ ～用途⇒技術の関係～

【分析目的】

自社で検討中の事業実現のために重要な解決技術や代替技術、競合他社の存在を把握し、事業化のための開発戦略や他社との協業戦略を検討する

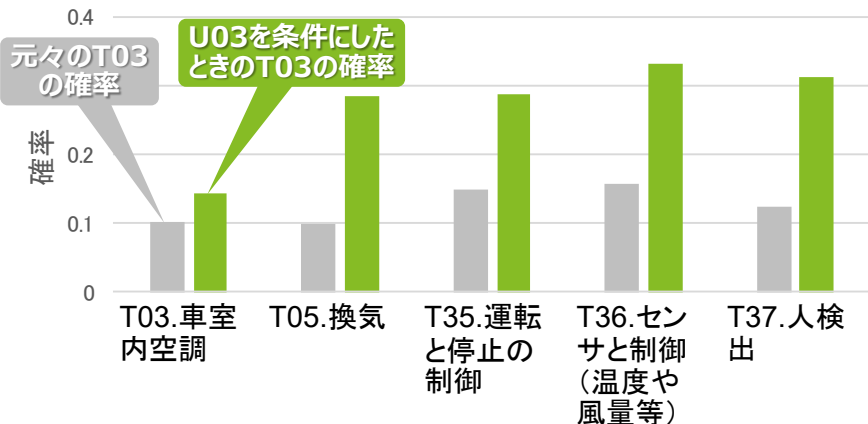
ベイジアンネットワークを適用して、用途トピックに対する技術トピックの確率的因果関係をモデル化します



用途と関係のある技術の確認

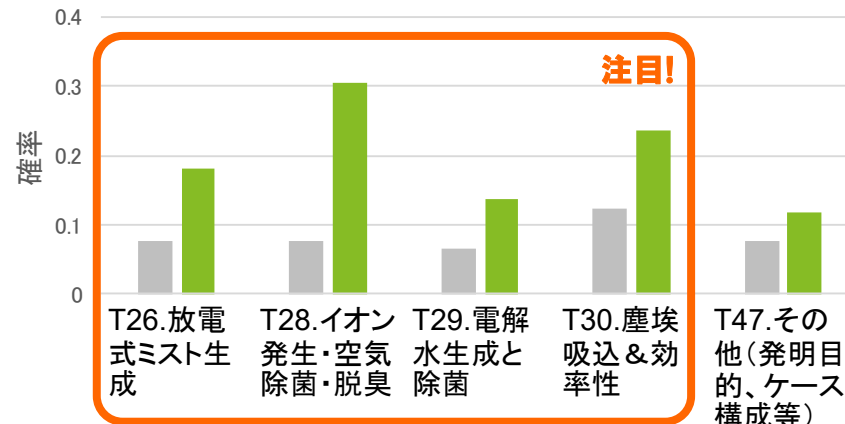
ベイジアンネットワークによって、1つの用途トピックを条件に与えたときの各技術トピックの確率の変化をシミュレーションし、用途に対する技術の関連性の強さを確認します

「U03.空調の省エネ、快適性」と関係のある技術



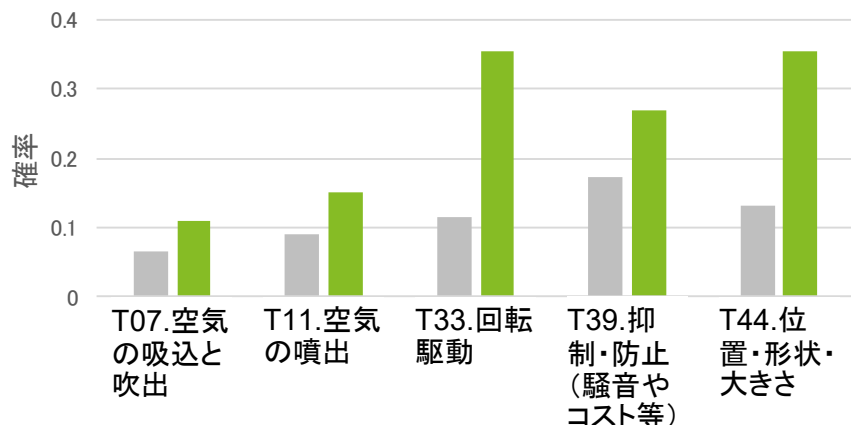
■ 元々の発生確率(事前確率) ■ U03を条件に与えた確率

「U06.空気浄化(除菌・脱臭)」と関係のある技術



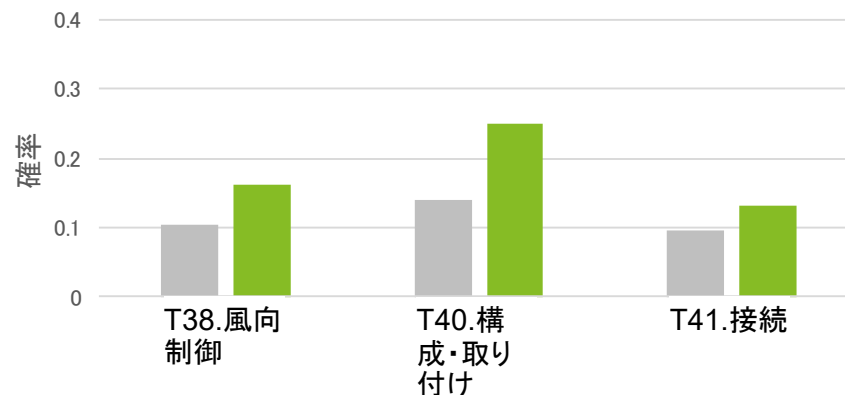
■ 元々の発生確率(事前確率) ■ U06を条件に与えた確率

「U15.騒音低減」と関係のある技術



■ 元々の発生確率(事前確率) ■ U15を条件に与えた確率

「U22.構造の簡素化」と関係のある技術

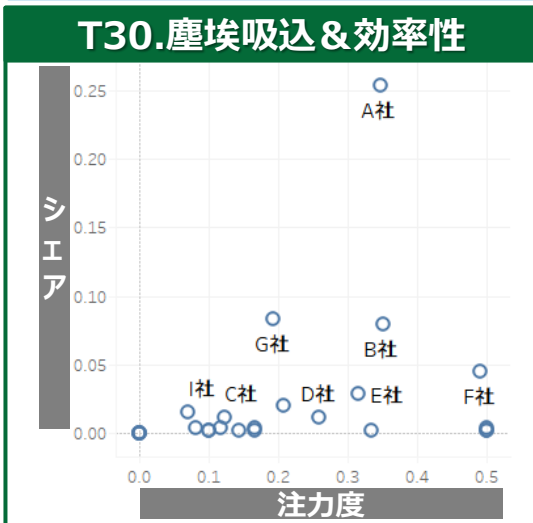
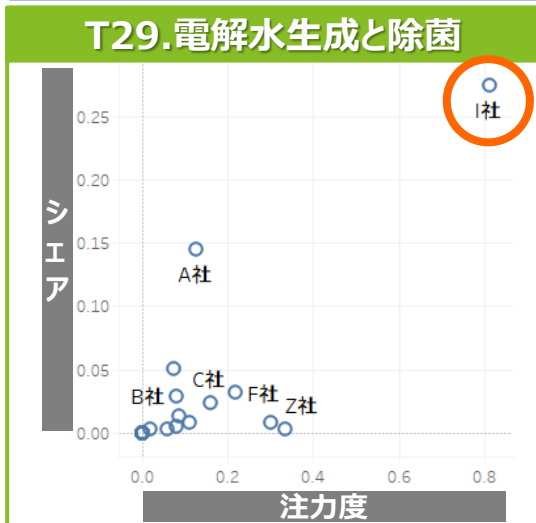
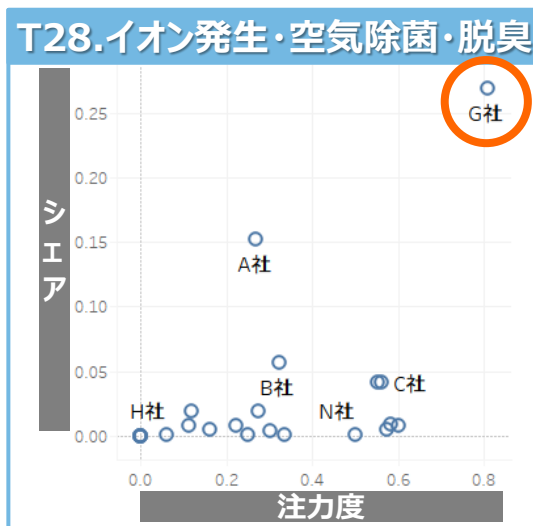
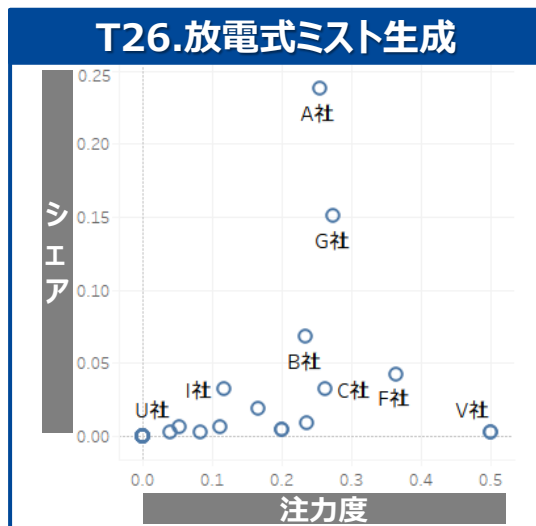


■ 元々の発生確率(事前確率) ■ U22を条件に与えた確率

用途「U06.空気浄化」と関係する技術トピックの出願人動向

U06の用途と関係する4つの技術のうち2つは一強状態にあり、U06の事業化では、この技術を避けた他の技術の開発を検討する、あるいはその一強企業の買収も考えられます

「U06.空気浄化」の関係技術トピックにおける出願人マップ



考察と戦略の検討

- 「T28.イオン発生・空気除菌・脱臭と「T29.電解水生成と除菌」は、それぞれG社とI社が高シェア高注力度のポジションを確立した一強状態の技術といえる
- 「T26.放電式ミスト生成」と「T30.塵埃吸込&効率性」は、シェアではA社が高く、注力度では例えばF社が高いが、高シェア高注力度の右上のポジションは空いている
- 一強状態の技術を避けて「U06.空気浄化」の用途を実現する場合、T26やT30の技術の開発が狙い目といえるが、シェアの高いA社や注力度の高いF社の動向は要注目である
- 一強状態にあるT28やT29の技術において、その一強企業と提携あるいはM&Aを実現すれば、その技術領域ごと獲得できる

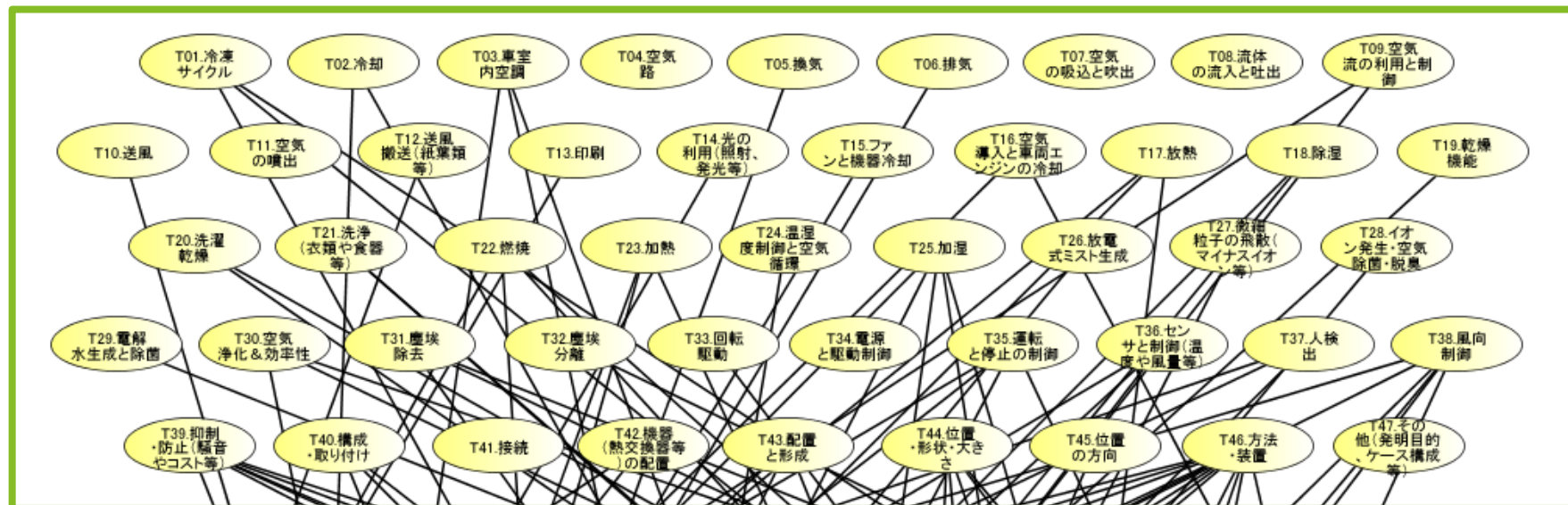
2. Nomolyticsを適用した特許分析事例

2-8. 用途×技術の関係分析＜その2＞ ～技術⇒用途の関係～

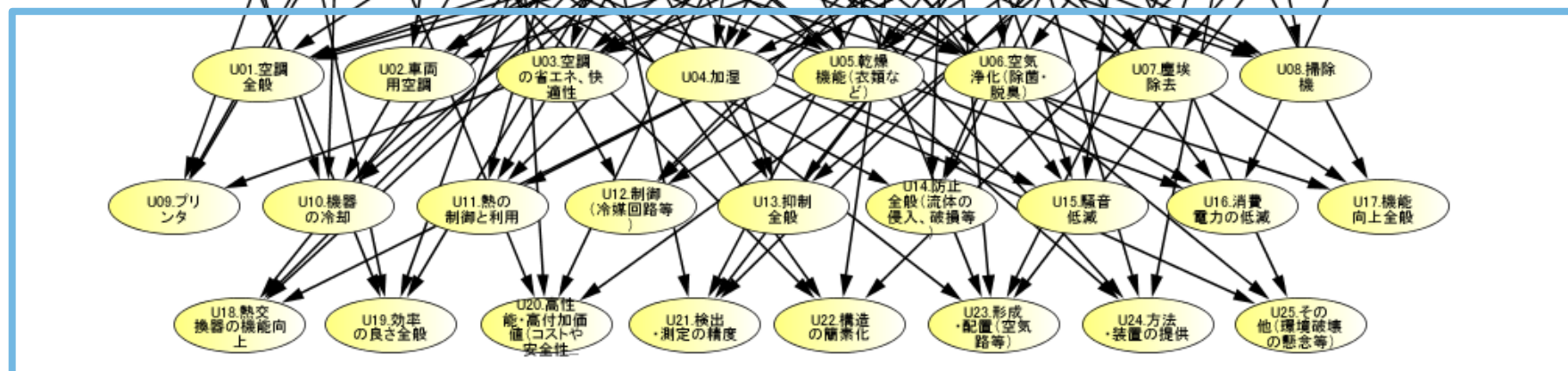
【分析目的】

自社技術と関係のある用途を把握し、まだ自社で想定していない用途を見つけ、保有技術を有効活用できる新しい用途展開のアイデアを創出する

ベイジアンネットワークを適用して、技術トピックに対する用途トピックの確率的因果関係をモデル化します



技術トピック

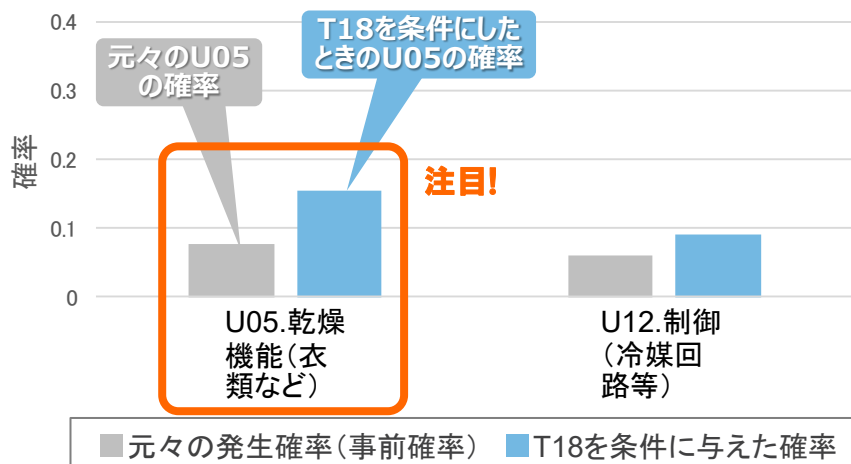


用途トピック

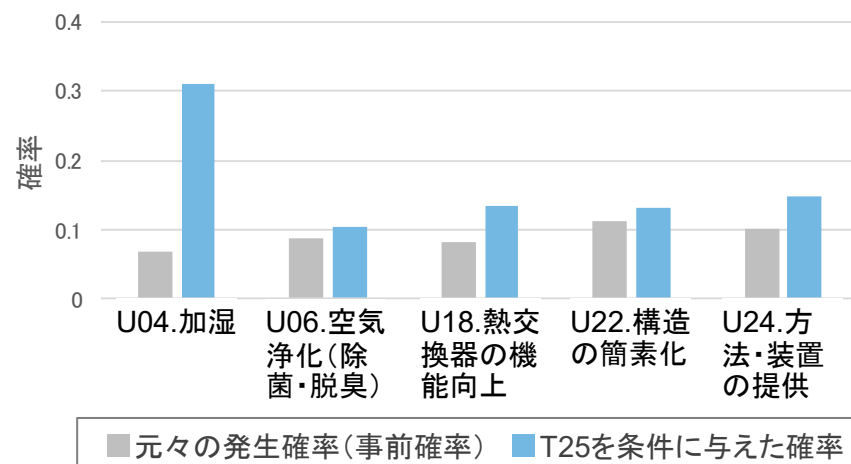
技術と関係のある用途の確認

ベイジアンネットワークによって、1つの技術トピックを条件に与えたときの各用途トピックの確率の変化をシミュレーションし、技術に対する用途の関連性の強さを確認します

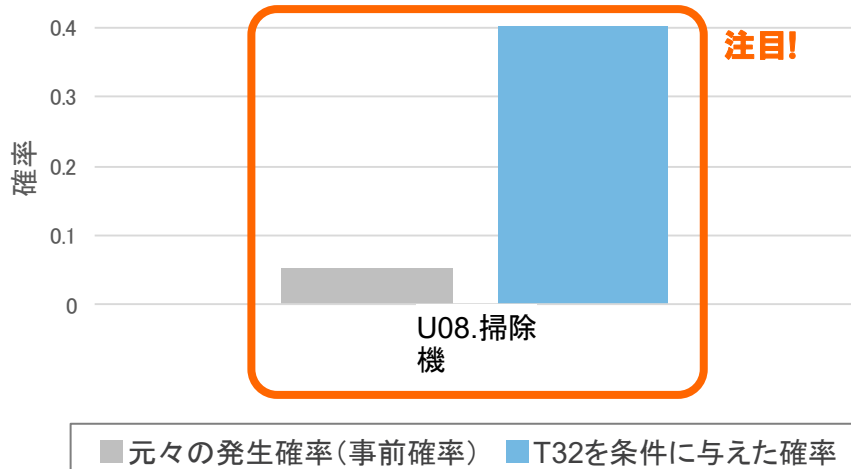
「T18.除湿」と関係のある用途



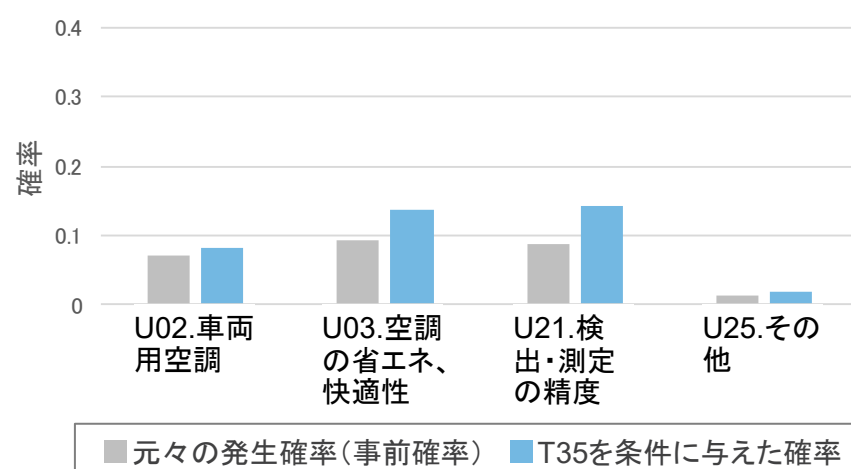
「T25.加湿」と関係のある用途



「T32.塵埃分離」と関係のある用途



「T35.運転と停止の制御」と関係のある用途



印刷機の中でインク液を吸収した用紙の湿気をムラなく取り除く除湿技術は、洗濯乾燥機の中で洗濯物をムラなく効率的に乾燥させることにも応用できるかもしれません

「乾燥機能」を想定した「除湿」の特許例

発明の名称

ドラム式洗濯乾燥機

【課題】

洗濯物を短い時間でムラ無く乾燥させ、乾燥工程の時間を短くすることができるドラム式洗濯乾燥機を提供する。

【解決手段】

送風機に吸い込まれた空気は、風路切替弁の切り替えにより、ドラム開口部に対向する前側吹出口へ流れたり、回転ドラムの後部に設けられた後側吹出口へ流れたりする。制御装置が風路切替弁の切り替えを制御することによって、恒率乾燥過程時、前側吹出口から乾燥用空気が吹き出し、かつ、減率乾燥過程時、後側吹出口から乾燥用空気が吹き出す。これにより、恒率乾燥過程において乾燥用空気が効果的に当たらなかった、回転ドラムの後端壁側の洗濯物に、乾燥用空気が減率乾燥過程で効果的に当たる。

「乾燥機能」を想定していない「除湿」の特許例

発明の名称

インクジェット記録装置及び画像記録方法

【課題】

処理液の厚みムラを低減するとともに処理液による用紙のコックリングを低減することで、高品質かつ高速の画像記録を可能とするインクジェット記録装置及び画像記録方法を提供する。

【解決手段】

記録媒体に処理液を付与する処理液付与部の後段には、記録媒体表面に残存する溶媒を蒸発させるプレ加熱部が設けられている。プレ加熱部はIRプレヒータにより記録媒体表面を輻射加熱するとともに、吸引ファンにより記録媒体表面の湿り空気を置換する。液状の処理液が不均一にならないように乾燥処理を施すことで、均一な膜厚を持つ固体状の凝集処理層が形成される。その後、本加熱部による熱風噴射加熱により、コックリング量が所定量以下になるように本加熱処理が施される。

※対外説明用のため要約文は一部加工している

印刷機でトナーを分離・回収するサイクロン部の清掃時期を判断して分離効率を維持する技術は、サイクロン掃除機の集塵部の集塵性能向上にも応用できるかもしれません

「掃除機」を想定した「塵埃分離」の特許例

発明の名称

電気掃除機

【課題】

集塵性能が向上しメンテナンスの軽減が図れる電気掃除機を提供すること。

【解決手段】

塵埃を含む空気を巡回させ塵埃分離する略円筒状の1次巡回室と、1次巡回室に連通した2次巡回室と、1次巡回室の下方に位置し塵埃を溜める集塵室と、塵埃を圧縮する圧縮板と、塵埃が流入する流入口を有し、圧縮板の底面の一部に突出部を流入口から見て集塵室の奥側に配設する構成としたことより、集塵室内に入った塵埃は、圧縮板の突出部に引っかかり動きが止められ、流れに乗って2次巡回室や1次巡回室側に戻ることが無いいため集塵性能が向上し、排気筒の詰まり防止によるメンテナンスの軽減を図ることができる。

「掃除機」を想定していない「塵埃分離」の特許例

発明の名称

画像形成装置

【課題】

サイクロン部の清掃時期を適正に判断して、トナーの分離効率の低下を抑制することが可能な画像形成装置を提供する。

【解決手段】

画像形成装置は、トナー含有空気からトナーを遠心分離するサイクロン部と、サイクロン部によって分離されたトナーを回収する回収部と、サイクロン部によってトナーが分離された空気を通過させ、残留トナーを捕集するフィルタ部と、空気を吸引する送風部と、フィルタの汚れを検知する汚れ検知センサが設けられたトナー捕集部を備え、汚れ検知センサで検知されたフィルタの汚れから推定した風量と、風速センサで取得した風量の実測値の差分が、サイクロン清掃閾値を超えたと判断すると、サイクロン部の清掃モードを実行する。

※対外説明用のため要約文は一部加工している

これまで培ってきた技術や経験と関連のある用途をいかに発想できるかということがイノベーションの鍵になります

サイクロン掃除機



ダイソンの吸引力が落ちないサイクロン掃除機は、製材工場の屋根にあった木くずと空気を分離するサイクロン装置をヒントに生まれた



サイクロン掃除機の技術はダイソンの様々な商品に応用されている

羽のない 扇風機



空気清浄 ファンヒーター



加湿器



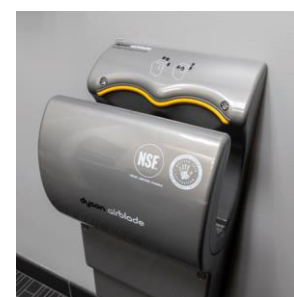
ヘアドライヤー



ヘアスタイラー



ハンドドライヤー



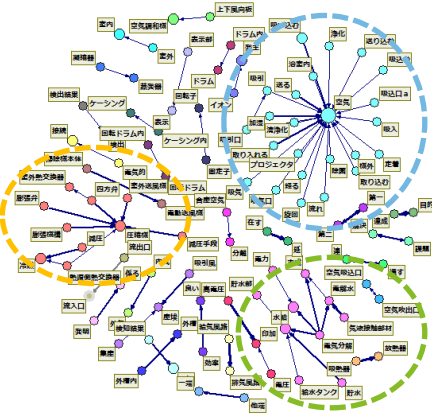
3. Nomolyticsを適用した特許分析のメリット

単語ではなく集約されたトピックをベースにした分析を実行することで、膨大な特許情報に潜む特徴を分かりやすく理解することができます

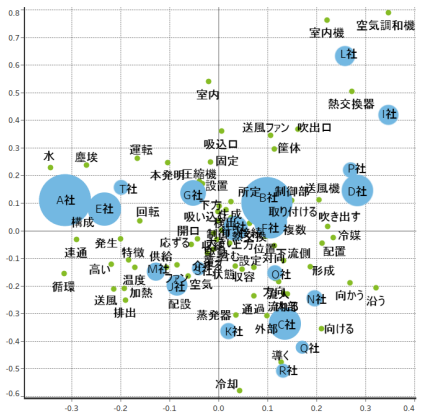
従来の特許分析

Nomolyticsを適用した特許分析

単語の共起ネットワーク



単語と出願人の対応マップ



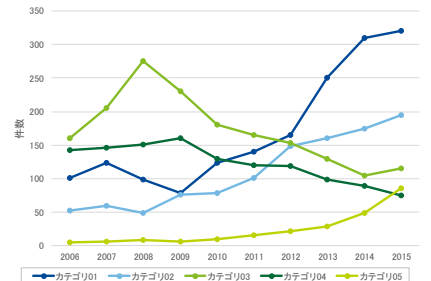
PLSAによって機械的に抽出されたトピック



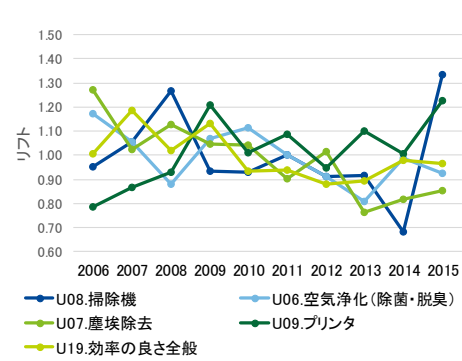
カテゴリのリスト作成

掃除機カテゴリのリスト	
掃除機	塵埃->分離
集塵	塵埃->吸い込む
集塵容器	塵埃->収容
吸引力	塵埃->遠心分離
サイクロン	含塵空気->分離

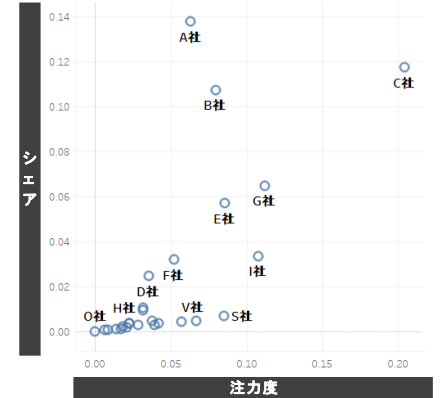
カテゴリ別のトレンド



トピック別のトレンド



トピック別の出願人の配置



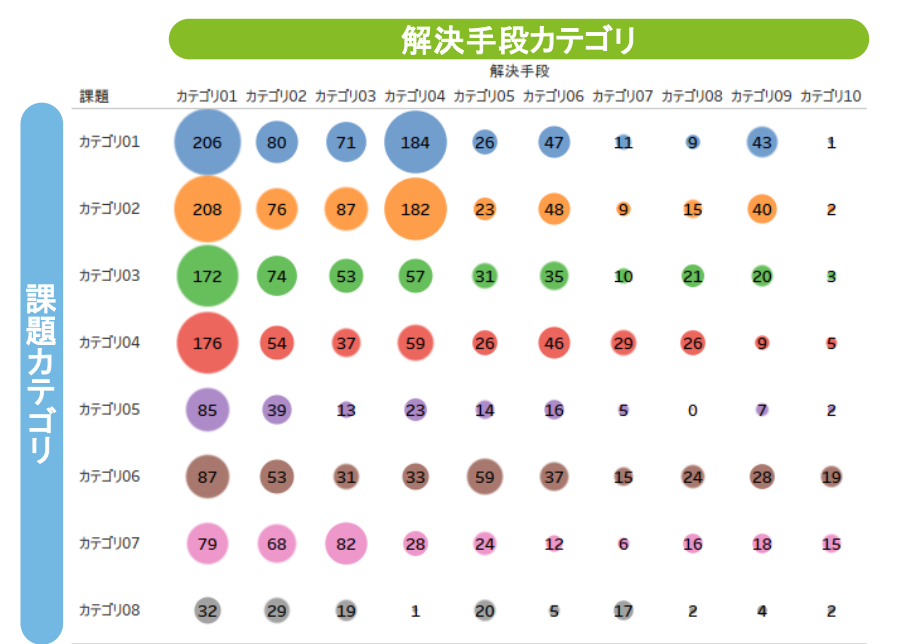
- 単語ベースの複雑なアウトプットから、文章全体に存在する話題を解釈したり、各出願人の特徴を把握しなければならない
- 単語を手でグルーピングしていくつかのカテゴリを作成し、カテゴリベースに分析するものの、そのカテゴリ作成は属人的で作業負荷も大きい

- 文章全体に存在する話題をPLSAで機械的に抽出できる
- 単語ではなくトピックをベースにトレンドや各出願人の特徴を分析し、分かりやすく理解することができる

用途と技術の統計的な関係を把握することで、用途を実現するための重要技術を確認して技術戦略を検討したり、自社技術を有効活用できる新規用途のアイデアを創出できます

従来の特許分析

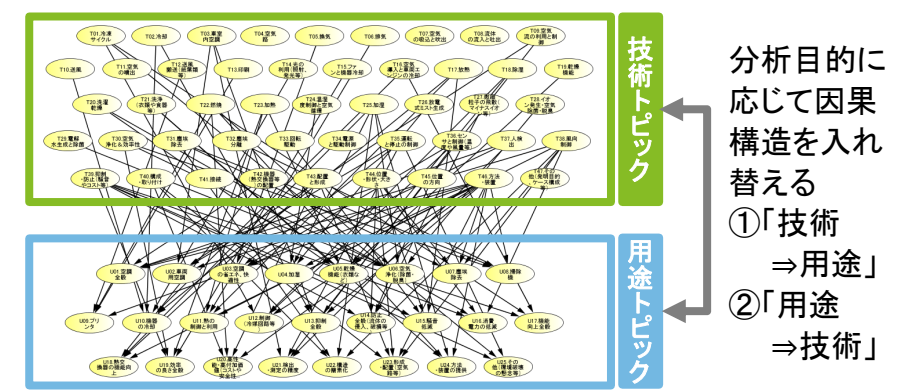
課題と解決手段のカテゴリ間のクロス集計



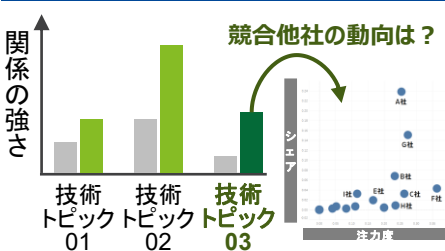
- 【課題】と【解決手段】それぞれに対して人がグルーピングして作成したカテゴリのクロス集計表を作成し、その対応関係を考察する
- その組み合わせで出願件数が多いからといって、統計的に意味のある関係であるとは限らない(全体的に出願件数が多いだけの可能性もある)

Nomolyticsを適用した特許分析

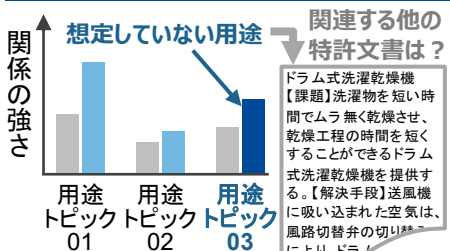
課題と解決手段のトピック間の統計的な因果関係モデル



想定用途と関係のある技術



自社技術と関係のある用途



- 客観的に抽出されたトピックをベースに課題と解決手段(用途と技術)の統計的な関係をベイジアンネットワークで把握できる
- 検討中の用途に対して、関係の強い技術を確認し、各技術における出願人の動向から自社の技術戦略を検討できる
- 自社技術と関係の強い用途で想定していないものを確認し、その関連特許の探索から技術の新規用途アイデアを創出できる

**資料に関するお問い合わせやコンサルティングの
ご相談は以下までお願いします。**

analytics.office@analyticsdlab.co.jp

**会社ホームページもご参考にしてください。
過去の講演・論文資料や技術解説も掲載しています。**

<https://www.analyticsdlab.co.jp/>

株式会社アナリティクスデザインラボ

