

現象理解の AI が繰り広げる新たな特許文書分析

トピックモデルの応用で探索する技術戦略のインサイト

株式会社アナリティクスデザインラボ
代表取締役 野守耕爾

2019年3月27日

人工知能技術を応用したデータ分析の研究開発とビジネスコンサルティングの経験を活かし、2017年6月にデータ活用コンサルティングの新会社を設立しました

株式会社アナリティクスデザインラボ

企業におけるデータ活用を支援するコンサルティング会社です。



データというスタートから課題の解決というゴールまでをいかにつなげばよいのか、どのようなデータ処理、分析手法、考察、アクションを検討していけばよいのか、というデータ活用するプロセスを企業の抱える課題や思惑・事情などに応じてしっかりとデザインし、それを実行することで企業の課題解決を支援します。

設立	2017年6月1日
事業内容	● 企業におけるデータ活用のコンサルティング ● データ分析技術の研究開発
資本金	5,000,000円
所在地	東京都中野区東中野1-58-8-204

野守 耕爾



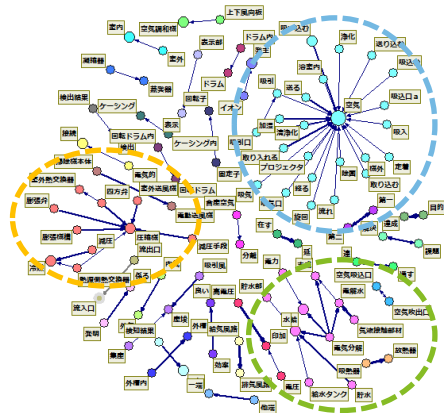
- 2012年3月
早稲田大学大学院 創造理工学研究科
経営システム工学専攻 博士課程修了
博士(工学)
 - 人間行動の計算モデルの開発を研究
- 2012年4月～(技術研修生としては2008年～)
独立行政法人産業技術総合研究所
デジタルヒューマン工学研究センター 入所
 - センシング技術を応用した子どもの行動計測と人工知能技術を応用した行動の確率モデルの開発を研究
- 2012年12月～
デロイトトーマツグループ 有限責任監査法人トーマツ
デロイトアナリティクス 入所
 - データサイエンティストとしてビッグデータを活用したビジネスコンサルティング及び分析技術の研究開発に従事
- 2017年6月～
株式会社アナリティクスデザインラボ 設立

人工知能技術を応用した新たな特許文書分析

これまでの特許文書分析

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、課題と解決手段の対応関係などを把握する分析がよく行われます

共起ネットワークによる全体像把握

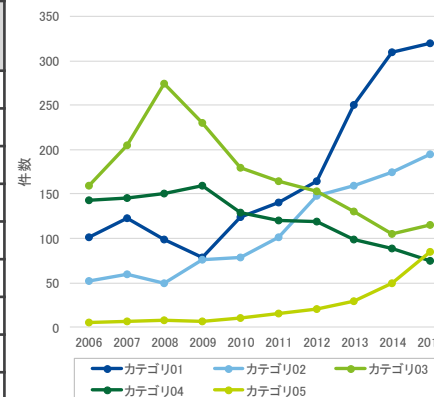


- 単語の共起関係をネットワークで可視化する
- ネットワークのかたまりを見ながら、全体でどのような話題が形成されているのか考察する

手動設定したカテゴリのトレンド把握

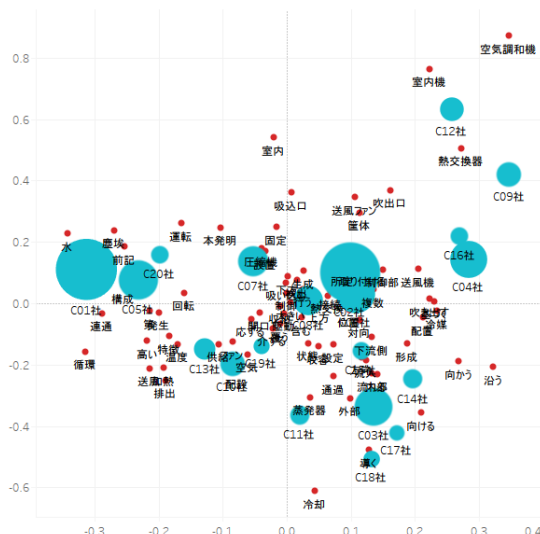
例) 掃除機カテゴリのリスト

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
- 各カテゴリの出願年ごとの出現頻度をグラフ化し、トレンドを把握する

コレスポネンス分析による出願人の特徴把握



- 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

課題と解決手段のクロス集計による関係把握

課題	解決手段									
	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08	カテゴリ09	カテゴリ10
カテゴリ01	206	80	71	184	26	47	11	9	43	1
カテゴリ02	208	76	87	182	23	48	9	15	40	2
カテゴリ03	172	74	53	57	31	35	10	21	20	3
カテゴリ04	176	54	37	59	26	46	29	26	9	5
カテゴリ05	85	39	13	23	14	16	5	0	7	2
カテゴリ06	87	53	31	33	59	37	15	24	28	19
カテゴリ07	79	68	82	28	24	12	6	16	18	15
カテゴリ08	32	29	19	1	20	5	17	2	4	2

- 「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

複数の人工知能技術を組み合わせることで、特許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

単語ベースの分析では
複雑で考察しにくい

カテゴリの設定が主観的で
作業負荷も大きい

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
人工知能技術

要因関係をモデリングする
人工知能技術

PLSA
確率的潜在意味解析

文脈を考慮した潜在的なトピック
(単語の集合)を抽出する

ベイジアンネットワーク

多様な要因間の確率統計的な
因果関係をモデル化する

PLSAは、トピックモデルと呼ばれる人工知能技術で、複雑なデータをいくつかの潜在変数で説明するクラスタリング手法として用いられます

PLSAの概要

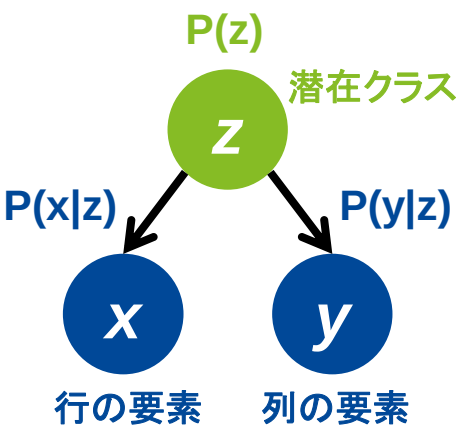
- 行列データの行の要素xと列の要素yの背後にある共通特徴となる潜在クラスzを抽出する手法である
- 元々は文書分類のための手法として開発されている (Hofman, 1999)
- 各文書の出現単語を記録した文書(行) × 単語(列) という高次元(列数の多い)共起行列データに適用することで複数の潜在トピックを抽出し、文書(行) × トピック(列) という低次元データに変換して文書を分類する

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	1		0	1
...						

文書ID	トピック1	トピック2	...	トピック11
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



- $P(z)$, $P(x|z)$, $P(y|z)$ 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

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

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

複雑な観測情報を分かりやすく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

- 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化する
- 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

特異値分解 LSA (Latent Semantic Analysis)

- $(m \times n)$ 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する)
- 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する)
- 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の拡張手法
- PLSA(他左3つの手法も含め)の過学習の問題に対して、LDAではディレクレ分布を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

ハードクラスタリング

- 一つの要素は必ず一つのクラスタに所属する
- 基本的に要素間の距離に基づいて分類を行う
- 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
- 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)

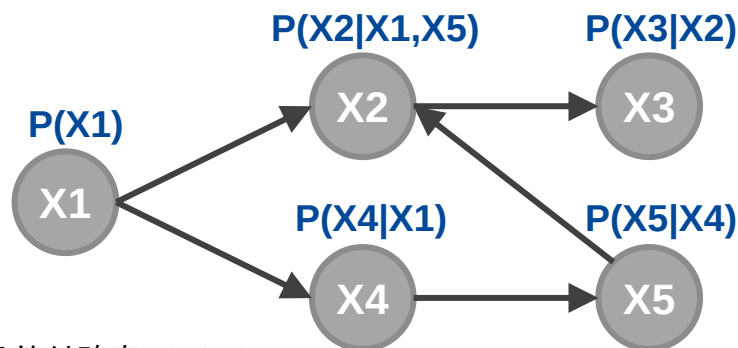
ソフトクラスタリング (潜在クラス分析、トピックモデル、次元圧縮)

- 一つの要素は全てのクラスに所属し、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
- 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
- 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

ベイジアンネットワークは、ベイズ推論に基づく人工知能技術で、変数間の確率的な因果関係を探索するモデリング手法として用いられます

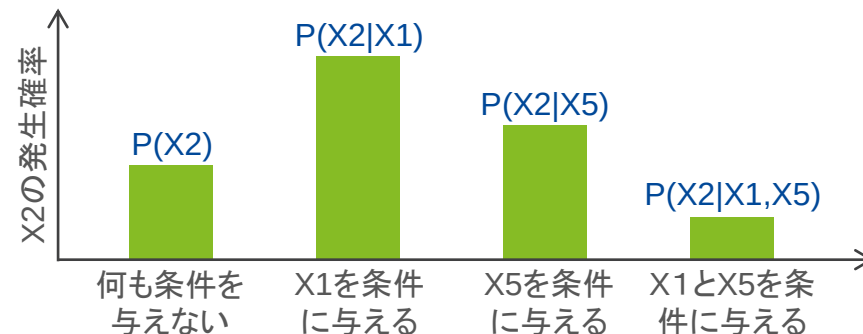
ベイジアンネットワークの概要

- 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

テキストで記された現象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)

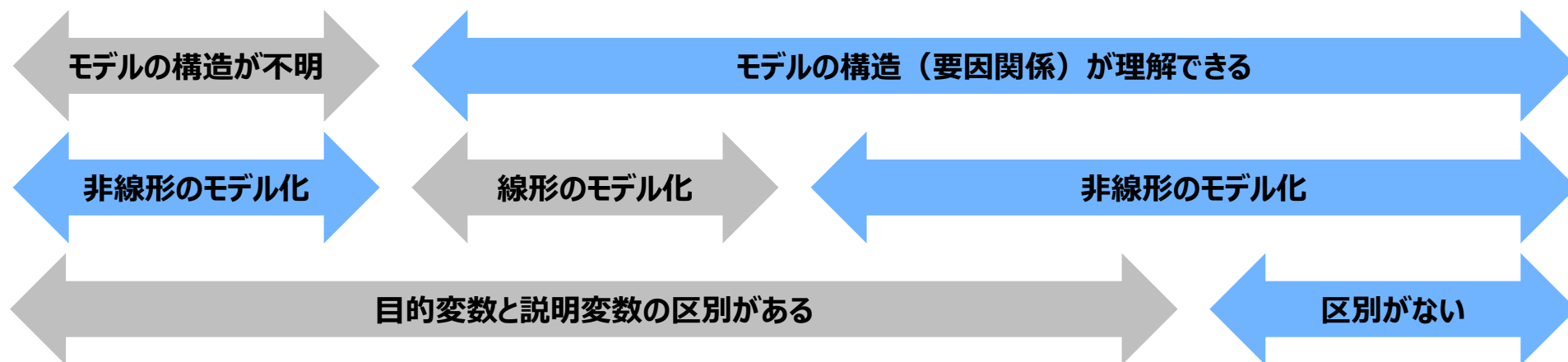
- 目的変数を説明変数の1次結合で定式化する
- 目的変数と説明変数の間に線形関係があるという仮定に基づいている
- 各説明変数の影響は独立しており、複合的な交互作用の影響は表現できない
- 説明変数間で相関が高い場合は解が不安定となり(多重共線性)、変数が多い場合この解消検討の負荷が大きい

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、複合条件によって効果が変わ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わる交互作用も表現できる



AI技術と一言で言っても様々あり、主に新規のデータを識別予測するAIと、今あるデータで記述された現象を理解するAIがあり、分析目的に応じて賢く使いこなすことが求められます

識別予測のAI

- ディープラーニング
- LDA
- 決定木
(バギング、ブースティング)

新規のデータに対してそれを
分類したり識別したりする



モデルの識別予測精度が命！

現象理解のAI

- ベイジアンネットワーク
- PLSA
- 決定木

現状のデータに潜む特徴
や要因関係を理解する



モデルのホワイトボックス化が命！

Nomolytics®: Narrative Orchestration Modeling Analytics

PLSA 確率の潜在意味解析

ベイジアンネットワーク

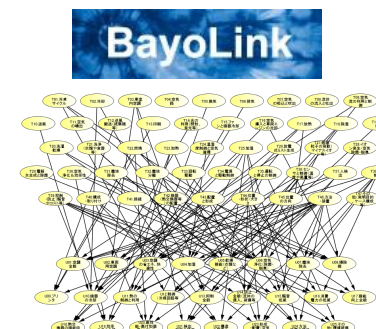
単語が出現する文脈を学習し、膨大な単語を複数のトピックにまとめる

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

トピック抽出

モデリング

Visual Mining Studio



テキスト情報内に潜
む複雑な要因関係を
構造化できる

条件を変化させたときの結果の挙動をシミュレーションできる

ある事象の発生確率
をコントロールする条
件を発見できる

特許要約の【課題】と【解決手段】から用途と技術のトピックを抽出し、トピックのトレンド分析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

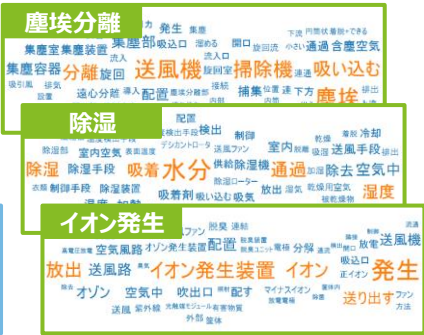
PLSAの適用

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
...	...	...

【課題】の要約文から用途トピックを抽出



【解決手段】の要約文から技術トピックを抽出



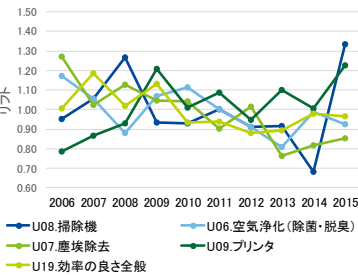
トピックのスコアリング

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	...	...	...	...	...	...	...	...

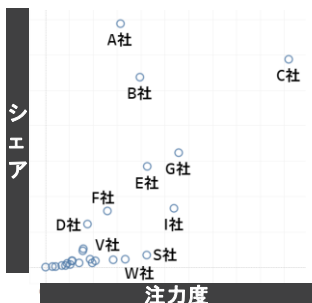
トレンドの分析

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



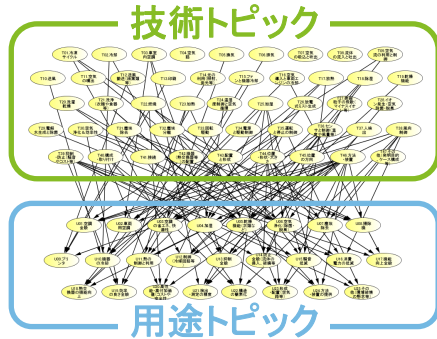
競合他社の分析

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



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出願動向を把握し、開発戦略や提携戦略を検討する

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

Nomolyticsを適用した特許分析事例

Nomolyticsを適用した特許分析事例

分析データ

「風」「空気」に関する10年分の特許データ30,039件を分析します

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～2015年

■ 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

【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。

Nomolyticsを適用した特許分析事例

トピックの抽出

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

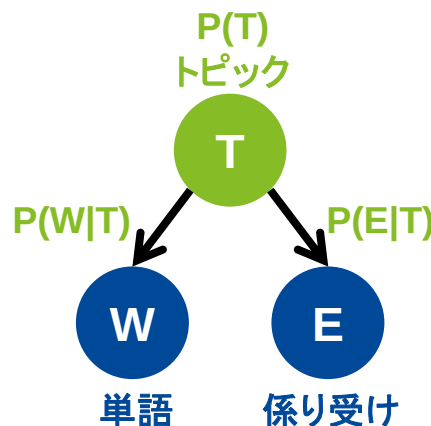
抽出した単語と係り受け表現に基づいて、「単語×係り受け表現」の共起行列(文章単位で同時に出現する頻度のクロス集計表)を作成する

	係り受け表現				
	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	...
単語					
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	
...					

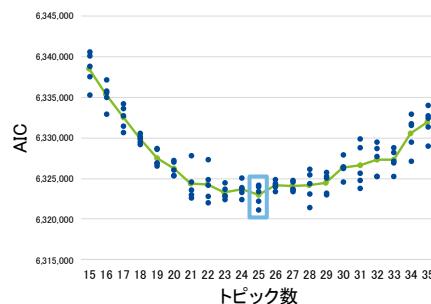
共起行列の構成(それぞれ頻度10件以上を対象)
 課題: 単語(3,256語)×係り受け(2,084表現)
 解決手段: 単語(5,187語)×係り受け(7,174表現)

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



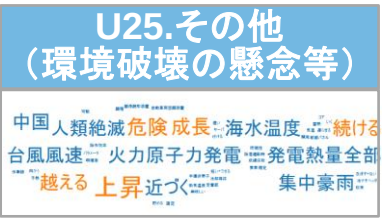
トピックの抽出

各トピックについて以下の3つの確率が計算される

- ① $P(T)$
トピックの存在確率
 - ② $P(W|T)$
トピックにおける単語の所属確率
 - ③ $P(E|T)$
トピックにおける係り受けの所属確率
- トピックにおける $P(W|T)$ と $P(E|T)$ からトピックの意味を解釈する

U04			
P(U)=2.6%			
P(W U)	単語	P(E U)	係り受け
5.5%	加湿装置	6.8%	加湿装置-提供
3.7%	水	3.1%	加湿器-提供
3.3%	供給	2.9%	ミスト発生装置-提供
2.4%	加湿	1.9%	水-供給
2.3%	カビ	1.7%	細菌-繁殖
2.1%	加湿器	1.5%	加湿-行う
2.1%	発生	1.4%	加湿機能付空気清浄機-提供
2.0%	繁殖	1.3%	ミスト-噴霧
1.9%	ミスト	1.3%	繁殖-抑制
1.7%	加湿性能	1.2%	十分-量
1.5%	ミスト発生装置	1.2%	カビ-発生
1.4%	細菌	1.2%	効率-良い
1.3%	室内	1.2%	空気調和機-提供
1.3%	抑制+できる	1.1%	加湿-加湿装置
1.1%	浴室	1.1%	空気-加湿
...	...	...	...

【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



【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**



20 特許、論文データ活用セミナー 20190327

Nomolyticsを適用した特許分析事例

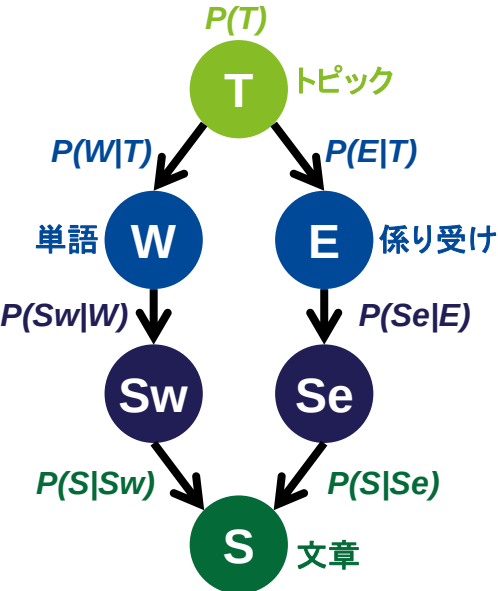
トピックのスコアリング

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{0:該当無,1:該当有}のデータに変換します

文章単位
のスコア

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

- 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Swと係り受けで定義される文章Se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 W_i で定義される文章 Sw_h
$Sw_h = \{W_1, W_2, \dots, W_i\}$
トピック T_k を条件とした文章 Sw_h の出現確率
$P(Sw_h T_k) = \sum_i P(Sw_h W_i)P(W_i T_k)$
単語 W_i が出現する中で文章 Sw_h が出現する確率(W_i の出現文章数の逆数)
$P(Sw_h W_i) = 1/n(W_i)$
係り受け E_j で定義される文章 Se_h
$Se_h = \{E_1, E_2, \dots, E_j\}$
トピック T_k を条件とした文章 Se_h の出現確率
$P(Se_h T_k) = \sum_j P(Se_h E_j)P(E_j T_k)$
係り受け E_j が出現する中で文章 Se_h が出現する確率(E_j の出現文章数の逆数)
$P(Se_h E_j) = 1/n(E_j)$
トピック T_k を条件とした文章 S_h の出現確率 ※ $P(S_h Sw_h)$ と $P(S_h Se_h)$ はともに1/2とする
$P(S_h T_k) = P(S_h Sw_h)P(Sw_h T_k) + P(S_h Se_h)P(Se_h T_k)$
文章 S_h の出現確率
$P(S_h) = \sum_k P(S_h T_k)P(T_k)$

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)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

Nomolyticsを適用した特許分析事例

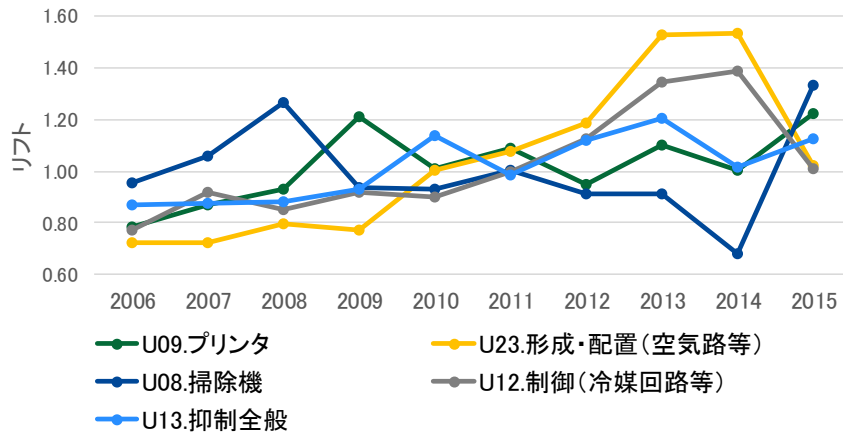
①出願年×トピックによるトレンド分析

【分析目的】

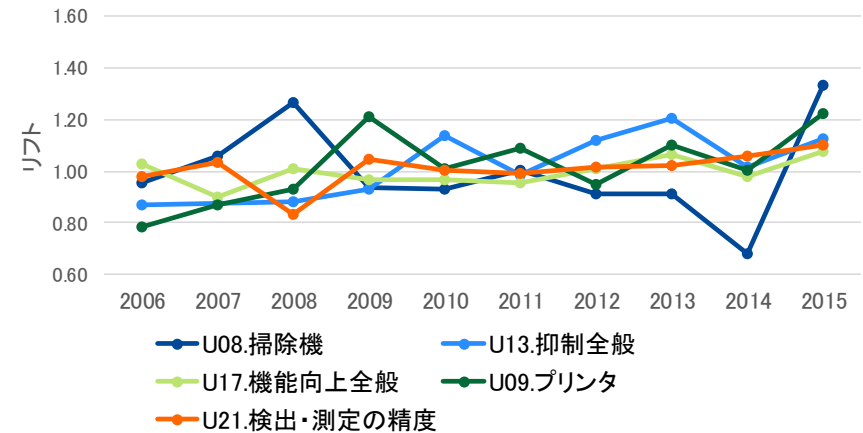
技術や用途のトレンドを把握し、有望なシーズやニーズを探り、今後の技術開発戦略を検討する

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

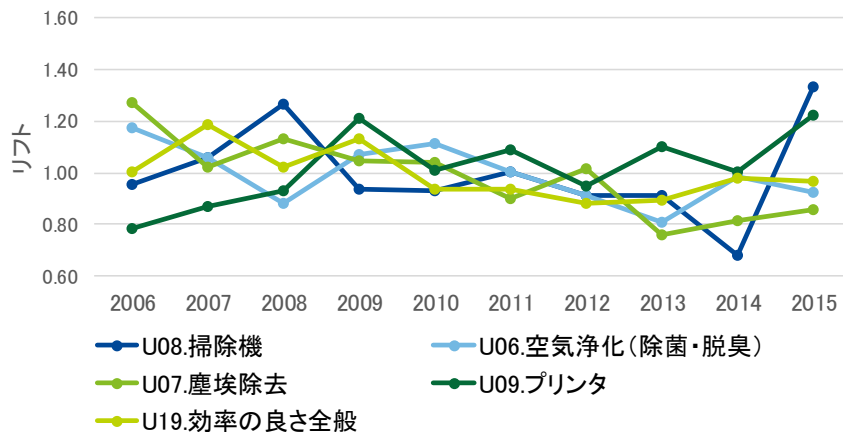
2006年からの上昇率 best5



2011年からの上昇率 best5



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

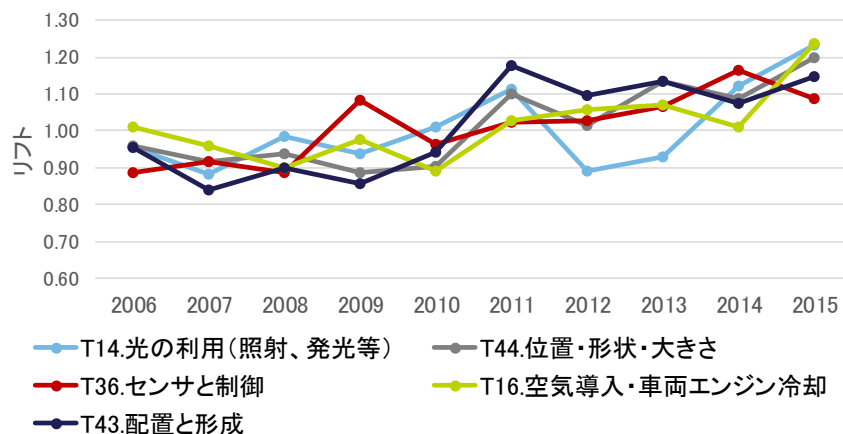
$$P(\text{出願年} \mid \text{トピック } U_x = 1)$$

$$P(\text{出願年})$$

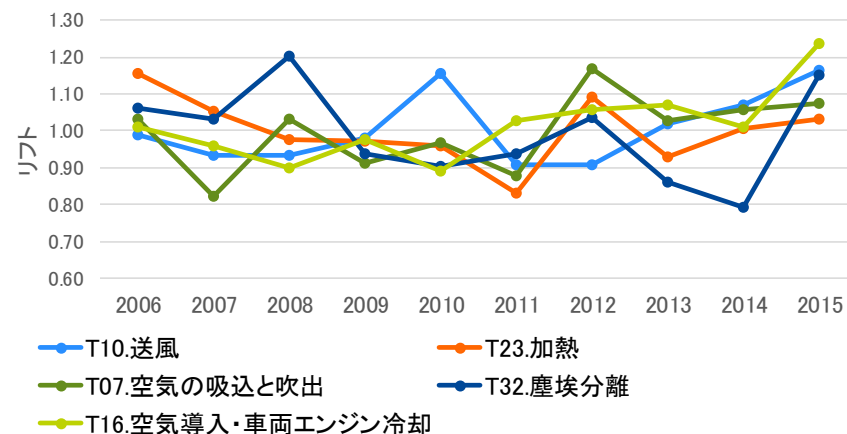
■ その出願年の出願件数割合を平均(=1)として標準化した値

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

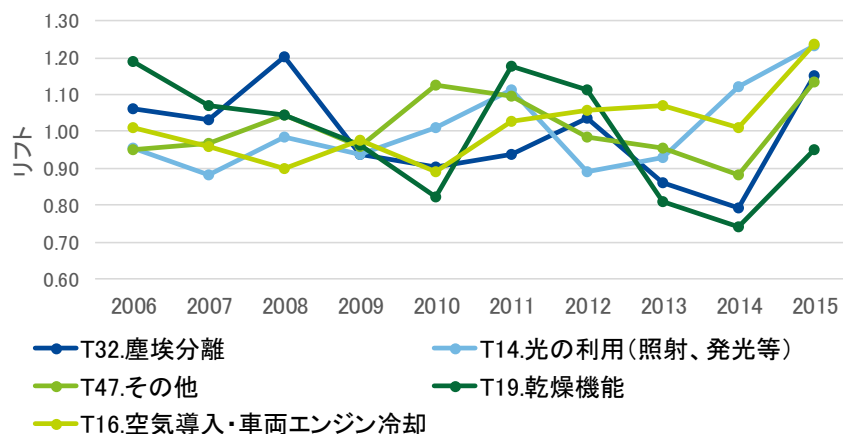
2006年からの上昇率 best5



2011年からの上昇率 best5



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

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

$$P(\text{出願年})$$

- その出願年の出願件数割合を平均(=1)として標準化した値

Nomolyticsを適用した特許分析事例

②出願人×トピックによる競合分析

【分析目的】

各出願人の動向を把握し、自社の技術開発戦略や差別化戦略、他社との提携戦略、自社技術の売却先などを検討する

注力度とシェアの2つの軸で各トピックにおける出願人の特徴を把握します

注力度(関心度)

- 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
- 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しており、高い技術を保有している可能性がある

$$P(\text{トピック}T = 1 \mid \text{出願人}X = 1)$$

シェア

- 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
- トピックのなかでどれくらいその出願人が占めているのかを示しており、技術市場のシェアともいえる

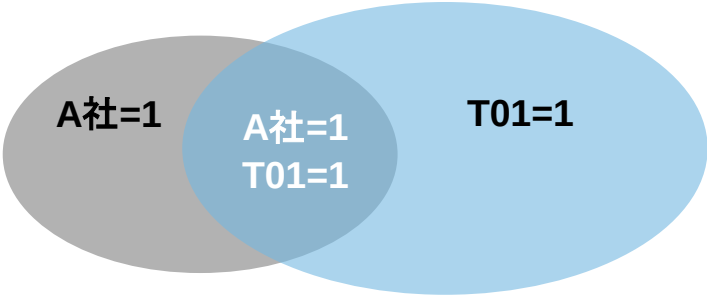
$$P(\text{出願人}X = 1 \mid \text{トピック}T = 1)$$

集計対象

- 名寄済みの出願人情報のうち、出願件数上位26社を集計対象とする

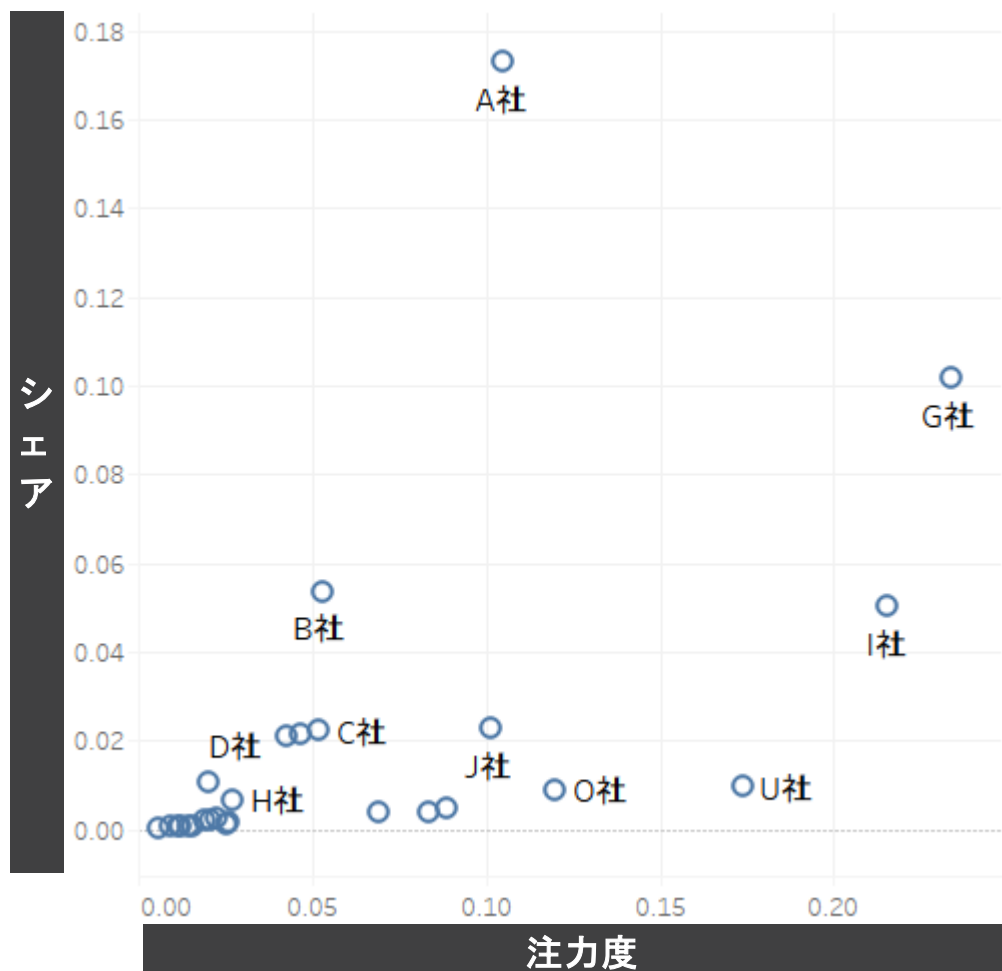
集計データと集計イメージ

特許ID	出願人 A社	出願人 B社	出願人 C社	...	トピック T01	トピック T02	トピック T03	...
1	1	0	0		1	0	0	
2	1	0	0		0	0	1	
3	1	0	0		1	0	0	
4	1	0	0		1	0	0	
5	1	0	0		1	0	0	
6	0	1	0		0	1	0	
7	0	1	0		0	1	1	
8	0	1	0		1	1	0	
9	0	1	1		0	0	0	
10	0	0	1		0	1	0	
...	...	...	...		...	...	...	



空気浄化に関する用途は、1社のシェアがとて高いものの注力度は高くなく、他に注力度のとても高い企業が存在するため、これらの企業と連携することで一強の存在になりえます

注力度とシェアの散布図

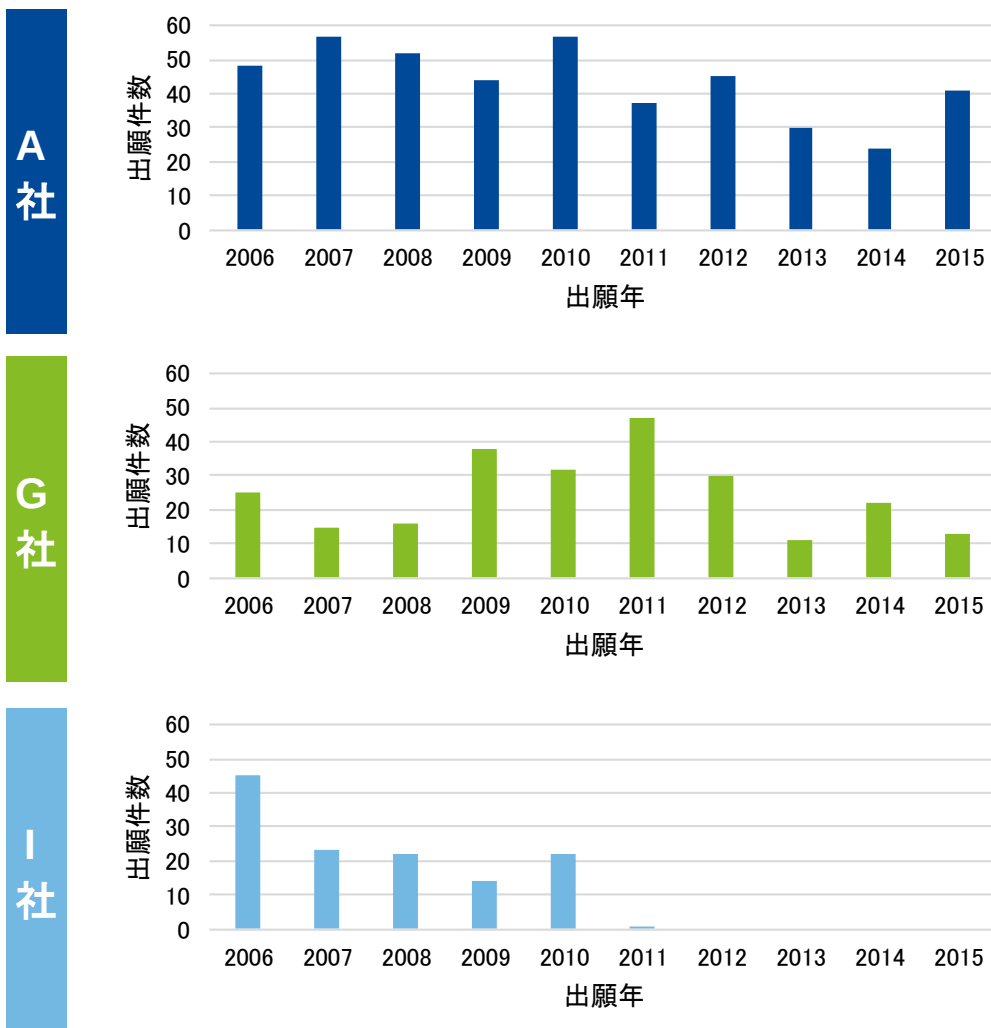


考察と戦略の検討

- A社は、シェアでは他社を大きく突き放しているが、注力度はそれほど高くはなく、A社にとって特に関心は強くない可能性がある
- G社とI社は注力度が高く、この用途に関心が高い出願人であり、この用途を実現するうえで高い技術を保有している可能性がある
- シェアではA社とのギャップはあるもののG社は第2位、I社は第4位のシェアがある
- A社は注力度の高いG社やI社と連携することで、高シェア高注力度のポジションを確立し、この用途において一強の存在となりうる
- G社とI社は、巨大なA社の関心が高くない間に独自の技術開発を進めることで、高い技術力による差別化を図ることができる
- あるいはG社とI社は、自社の独自技術を巨大なA社にライセンスや売却して収益化することも考えられる

注力度は高くないと思われたシェア1位のA社は、近年は出願件数を伸ばしており、今後積極的な開発や提携も考えられ、一方注力度の高いと思われたG社は近年出願が減っています

注目企業の出願件数のトレンド

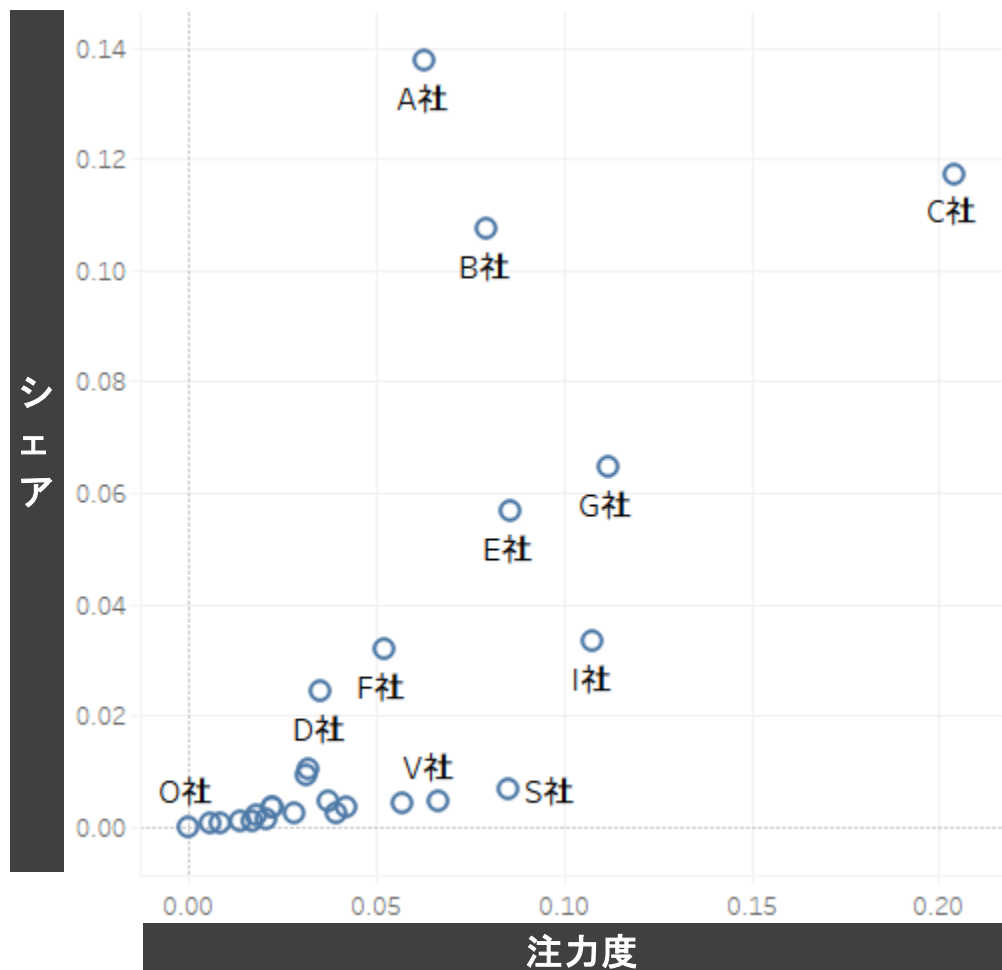


考察と戦略の検討

- シェア第1位のA社は、全体では注力度は高くなかったが、近年は出願件数を伸ばしており、この用途において関心が高まっている可能性がある
 - 今後積極的な技術開発や提携が考えられる
- 注力度第1位、シェア第2位のG社は、以前はこの用途の技術に注力して開発をしていた可能性があるが、近年は出願件数は減っており、現在はこの用途の関心は低くなっている可能性がある
 - 今後この用途に注力しないのであれば、これまで培ってきた技術を他社にライセンスや売却して収益化するという事も考えられる
- 注力度第2位、シェア第4位のI社は、2012年以降の出願件数はなく、この用途の技術開発を終了している、あるいは会社が存在しない可能性がある

塵埃分離に関する技術は、1社の注力度が高いものの、他にもある程度のシェア・注力度を保有する企業が何社か存在するため、今後連携などの動きも考えられる領域と思われます

注力度とシェアの散布図

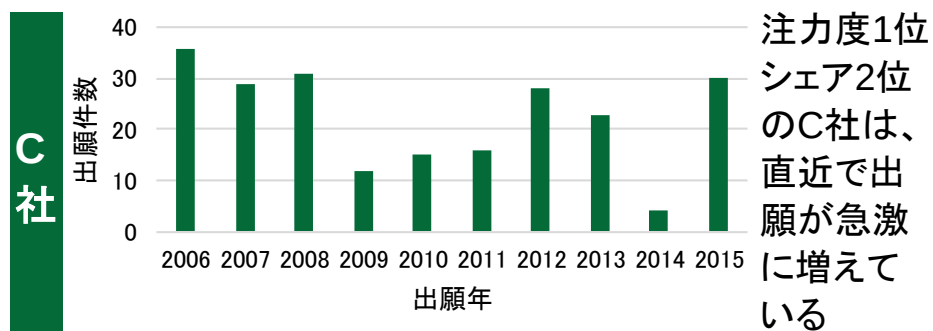
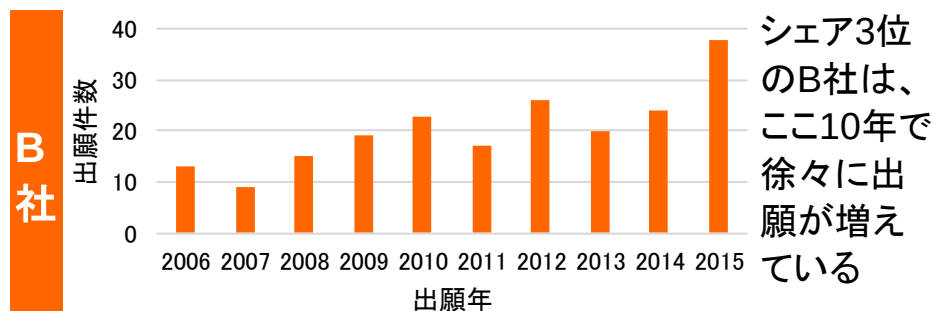


考察と戦略の検討

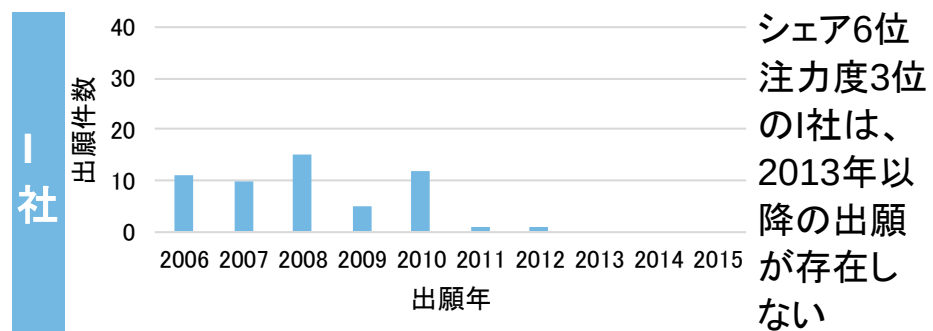
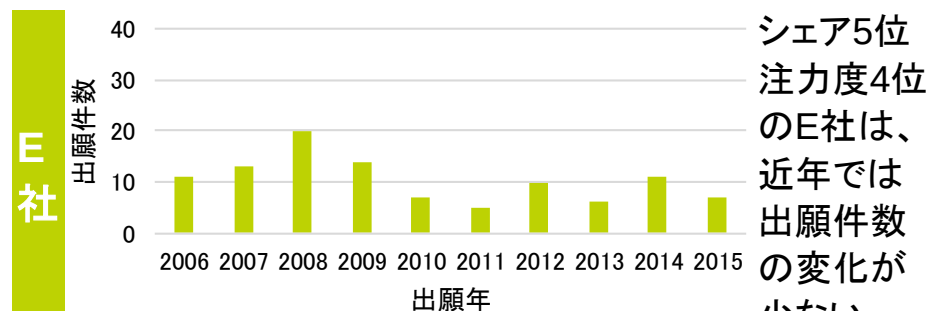
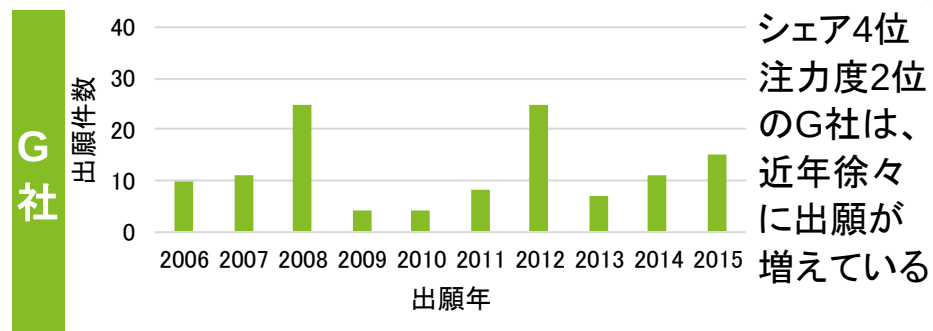
- C社は、高めのシェアを獲得しつつ、他社と比べて注力度がとても高く、高い技術力を保有していると考えられ、今後はよりシェアを伸ばすことで高シェア高注力度のポジションを確立することができると思われる
- A社とB社は、シェアは高いがまだC社に注力度で劣っているので、例えば規模は中程度だが注力度は比較的高く、技術力があると思われるE社、G社、I社などと連携することで、C社の上のポジションを狙うことができる可能性がある

注力度1位のC社は直近で出願が急増し、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社は増加で、またシェア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トレンド



注目企業の出願件数のトレンド



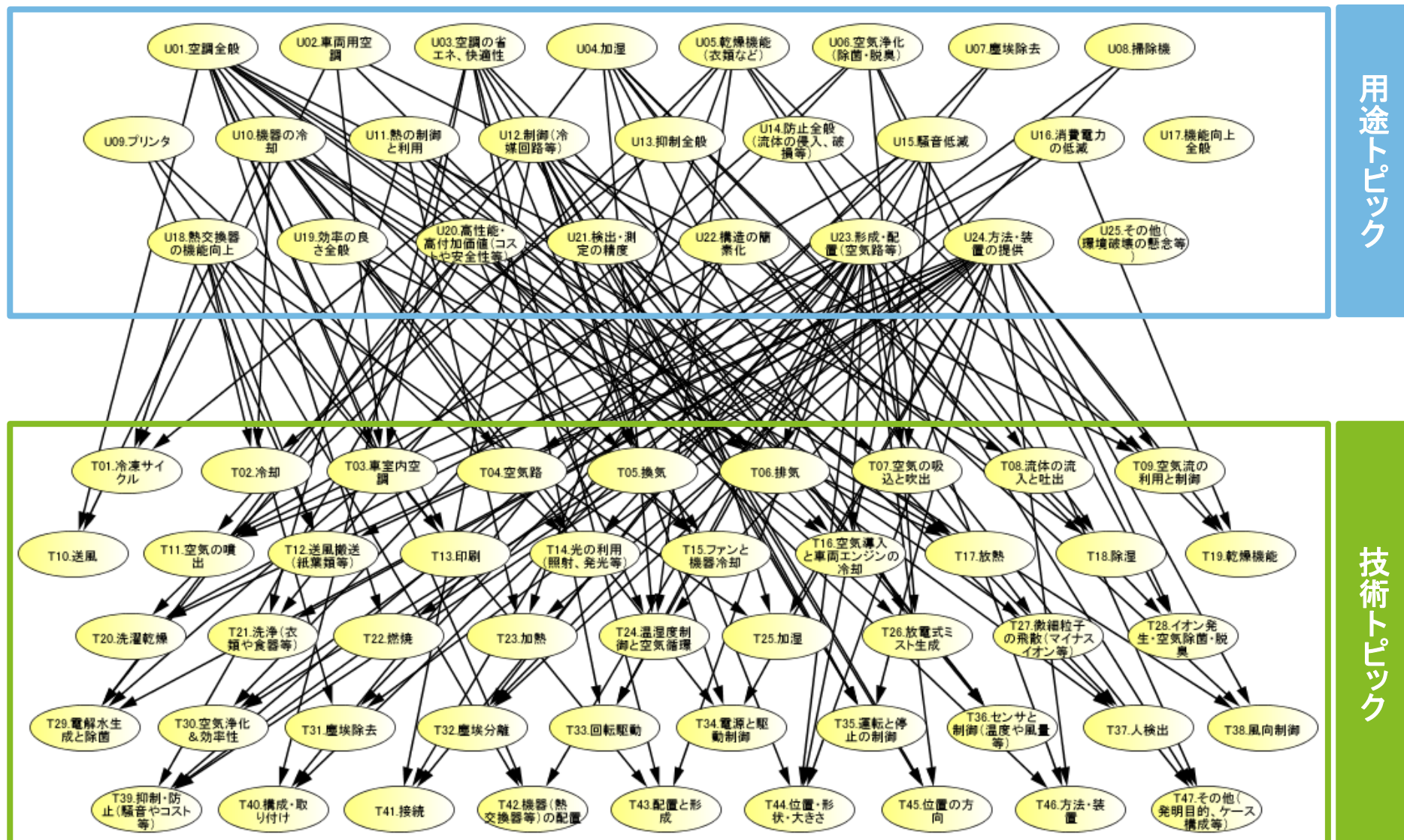
Nomolyticsを適用した特許分析事例

③用途トピック×技術トピックの関係分析 ＜その1＞用途⇒技術の関係

【分析目的】

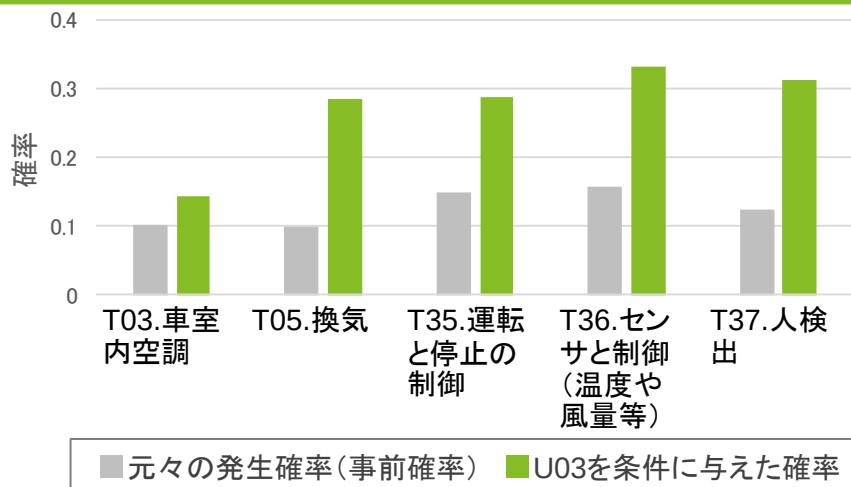
自社で検討中の用途に対する解決技術を把握し、その用途を達成するためにどの技術開発に注力すべきか、競合となりそうな他社はどこか、他社が牛耳る技術の代替技術は何か、どの出願人と連携すべきかなど、開発戦略や提携戦略を検討する

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

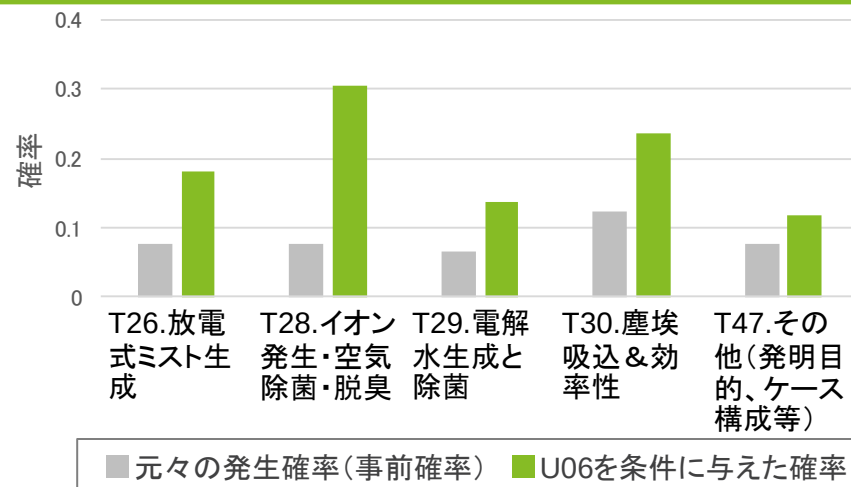


用途トピックを条件に与えたとき、それと確率的因果関係を持つと判定された各技術トピックの確率がどのように変化するのかシミュレーションして、その関連性の強さを確認します

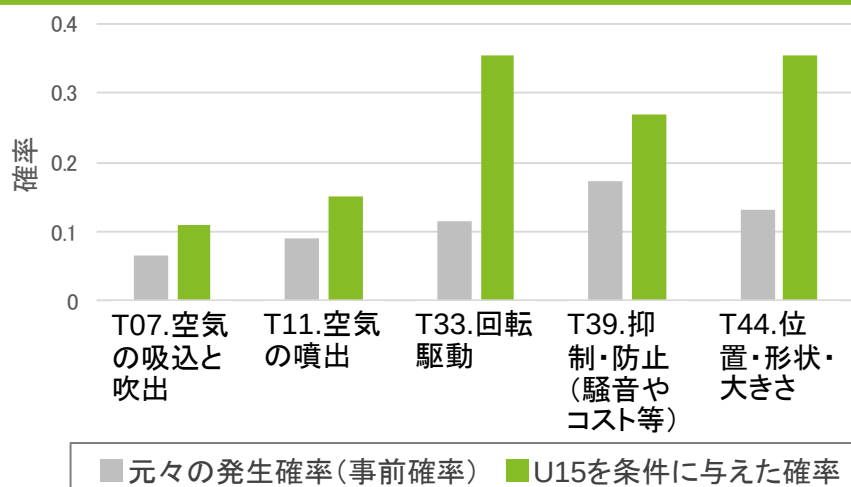
「U03.空調の省エネ、快適性」と関係のある技術



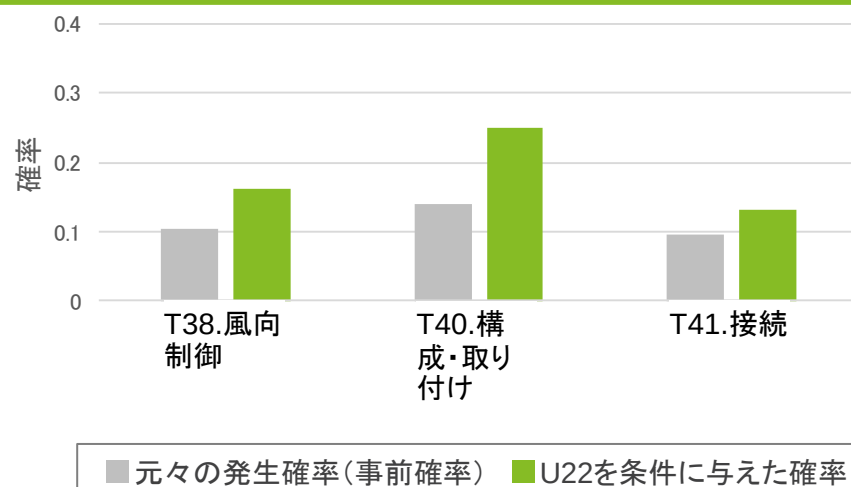
「U06.空気浄化(除菌・脱臭)」と関係のある技術



「U15.騒音低減」と関係のある技術

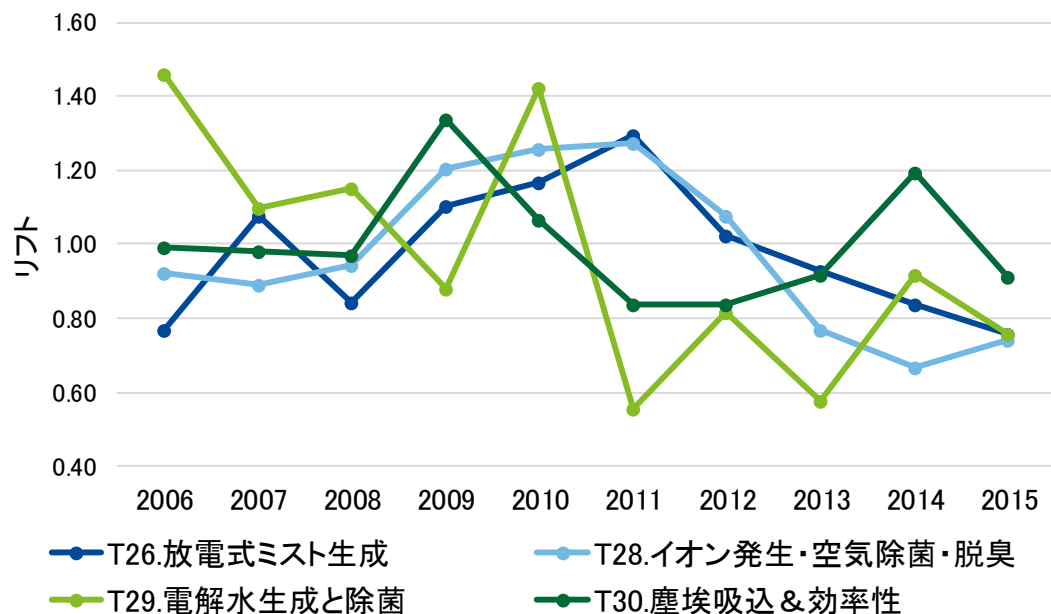


「U22.構造の簡素化」と関係のある技術



「U06.空気浄化」の用途に関する技術トピックでは、「T29.電解水生成と除菌」と「T30.塵埃吸込 & 効率性」が近年緩やかな上昇傾向にあります

「U06.空気浄化(除菌・脱臭)」の関係技術トピックのトレンド



考察

- 「T26.放電式ミスト生成」と「T28.イオン発生・空気除菌・脱臭」は2009年～2011年をピークに近年は下降傾向にある
- 「T29.電解水生成と除菌」と「T30.塵埃吸込 & 効率性」は2011年～2013年をボトムに近年はやや上昇傾向にある

集計の仕方

- リフト値を出願年・技術トピックごとに集計

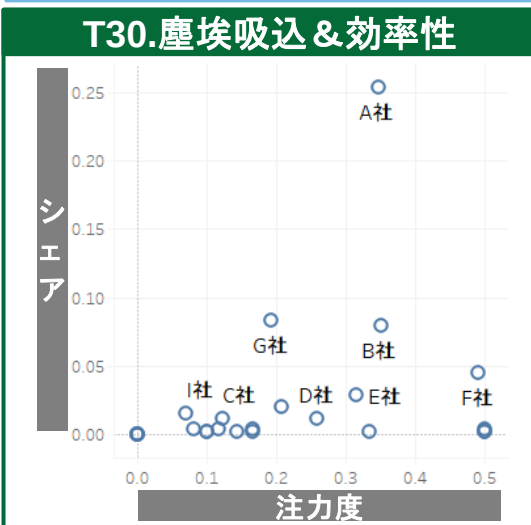
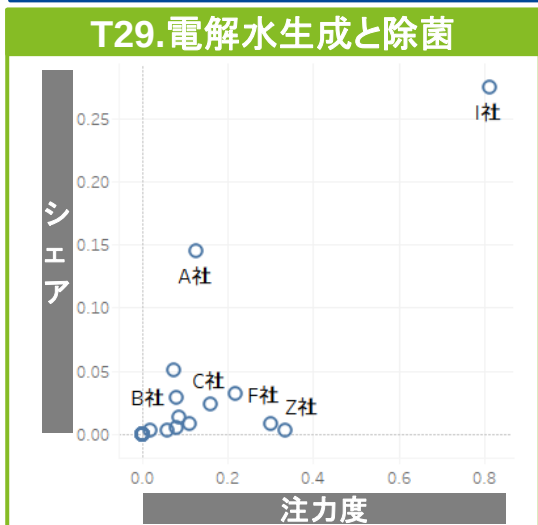
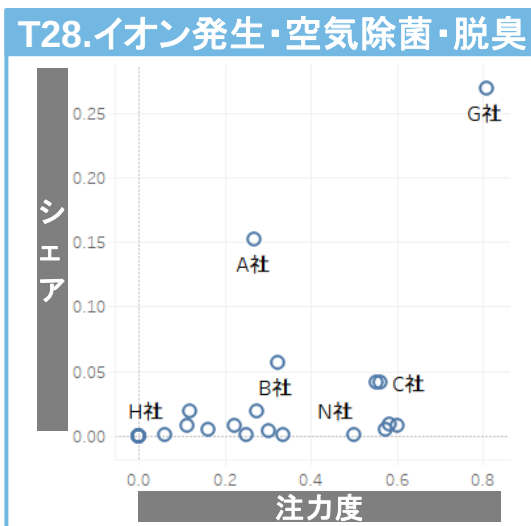
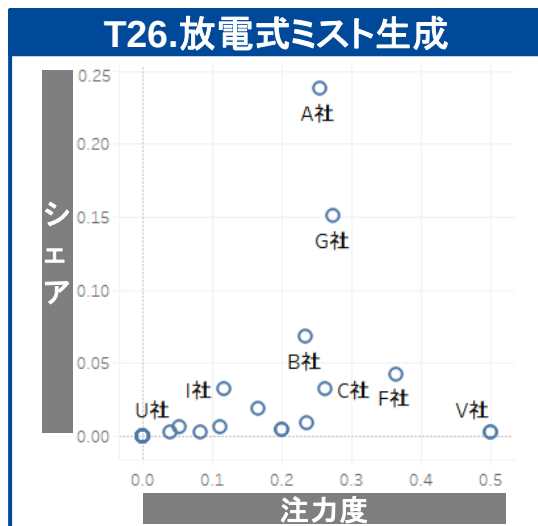
$$P(\text{技術トピック}T=1 \mid U06=1, \text{出願年}Y=1)$$

$$P(\text{技術トピック}T=1 \mid U06=1)$$

- U06における各技術トピックの該当割合を全体で計算した値(分母)は、各出願年条件下(分子)では何倍になるのかを示す
- U06における各技術トピックの該当割合を全体で計算した値を平均(=1)として標準化したもの

U06の用途と関係する4つの技術のうち2つは一強状態で、U06にアプローチするためには、この技術を避けた他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、高シェア高注力度のポジションは空いているが、どちらもF社がシェアは低いものの注力度が高いポジションにある
- 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の技術が狙い目といえるが、注力度の高いF社は要注目である
- 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

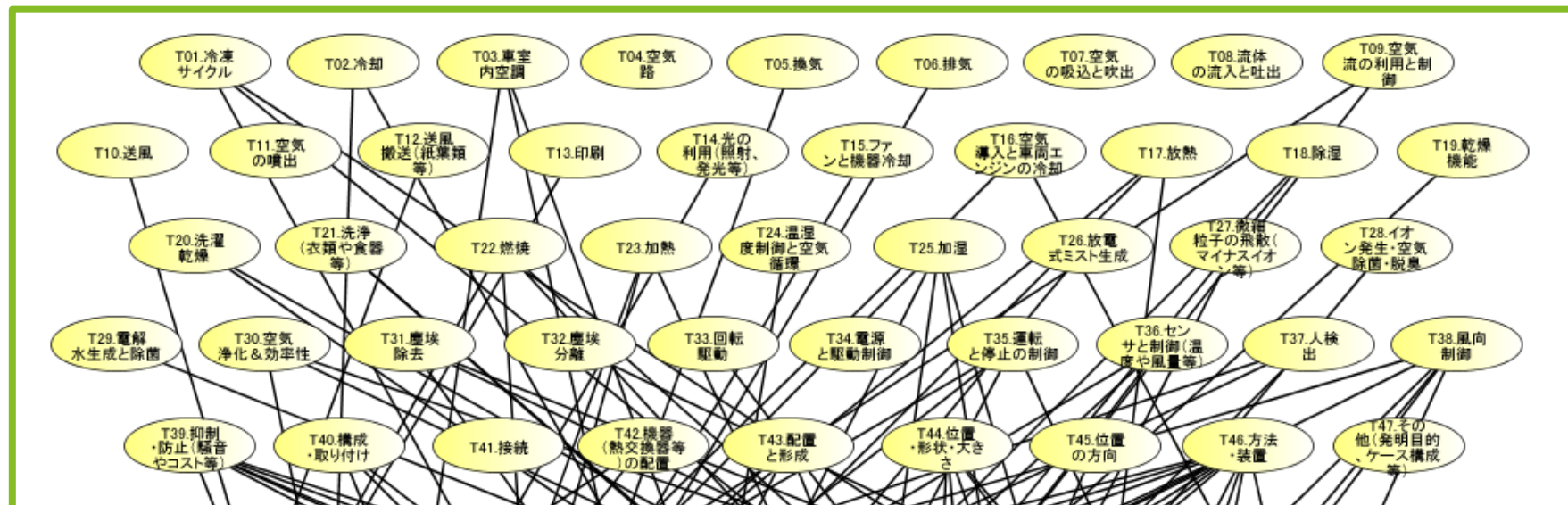
Nomolyticsを適用した特許分析事例

③用途トピック×技術トピックの関係分析 ＜その2＞技術⇒用途の関係

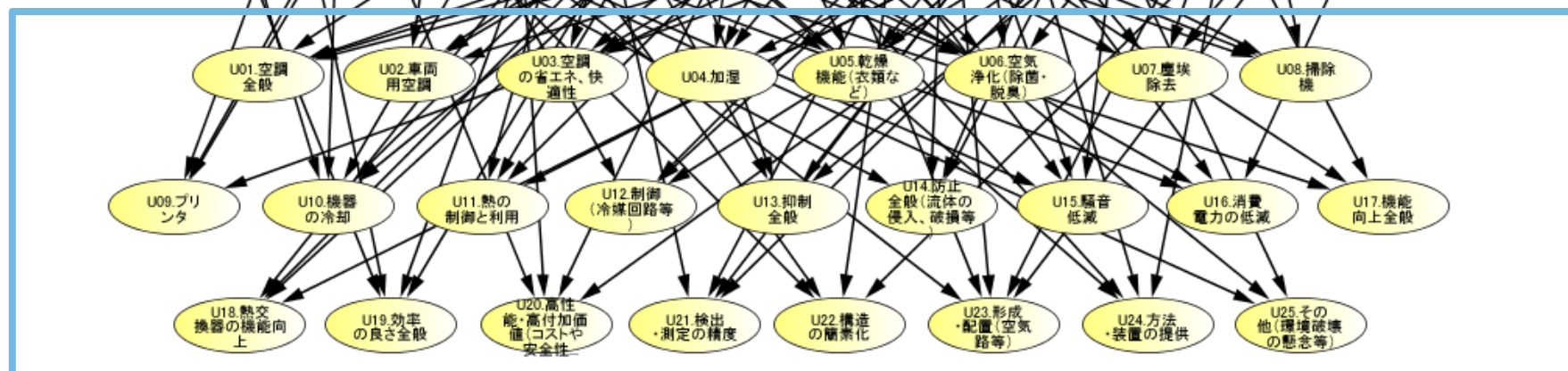
【分析目的】

自社技術と関係のある用途を把握し、そのうち自社で想定していない用途を見つけ、その用途に関連する他社の特許文書を探索することで、自社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



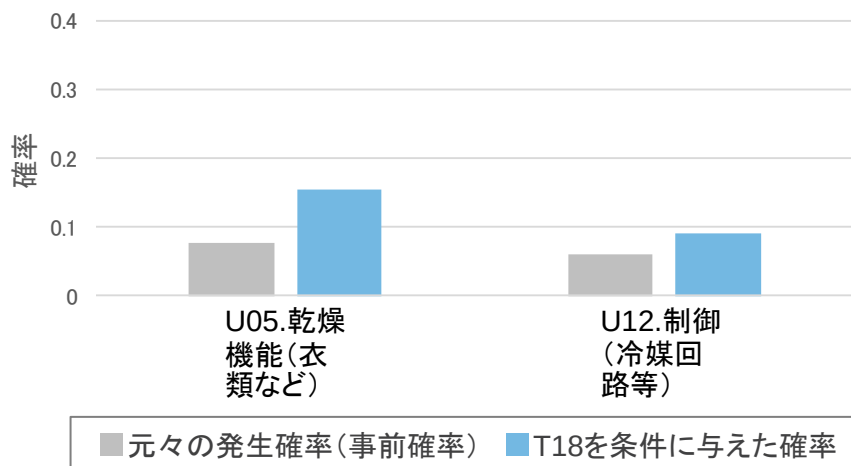
技術トピック



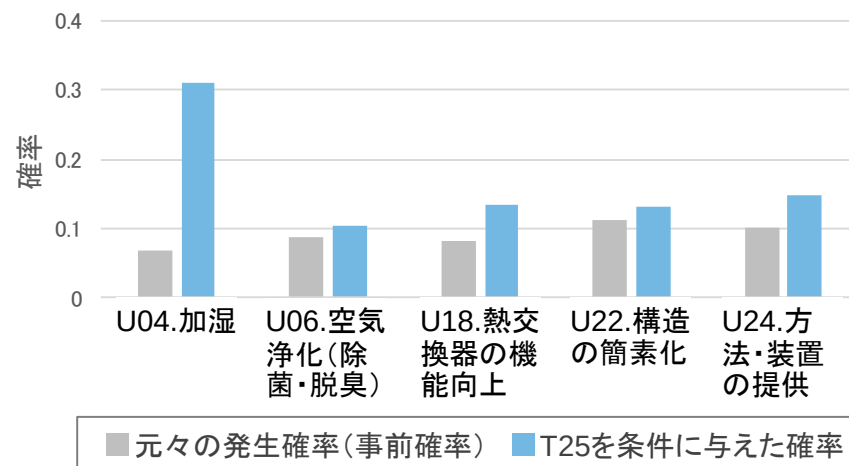
用途トピック

技術トピックを条件に与えたとき、それと確率的因果関係を持つと判定された各用途トピックの確率がどのように変化するのかシミュレーションして、その関連性の強さを確認します

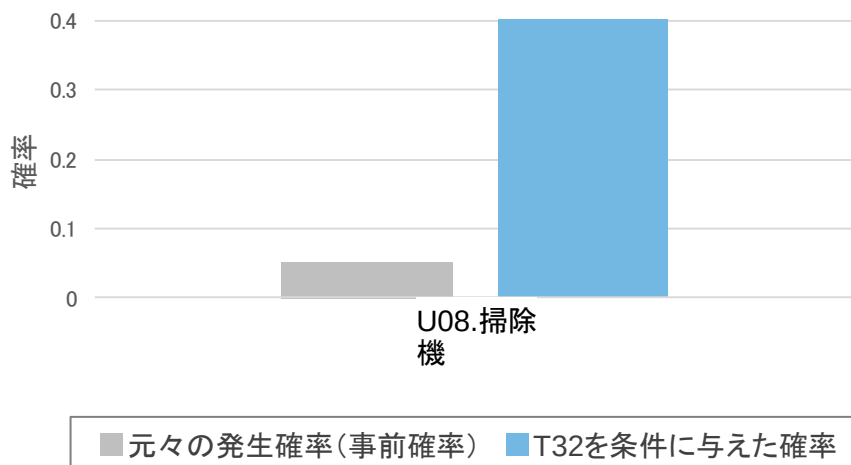
「T18.除湿」と関係のある用途



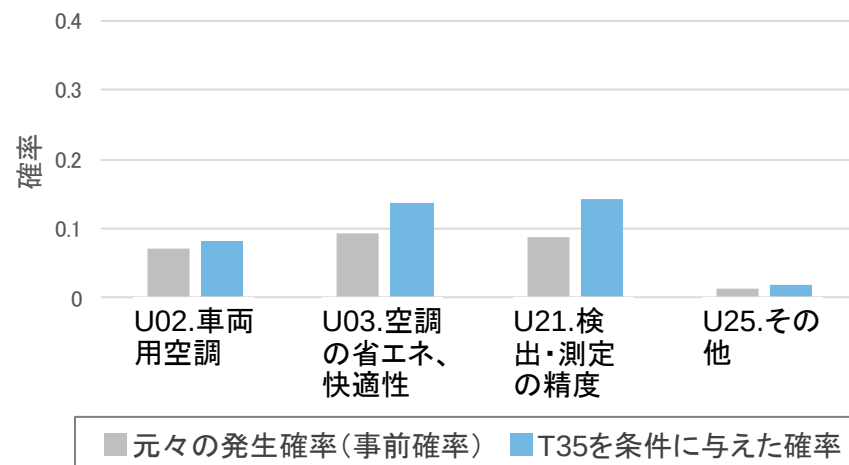
「T25.加湿」と関係のある用途



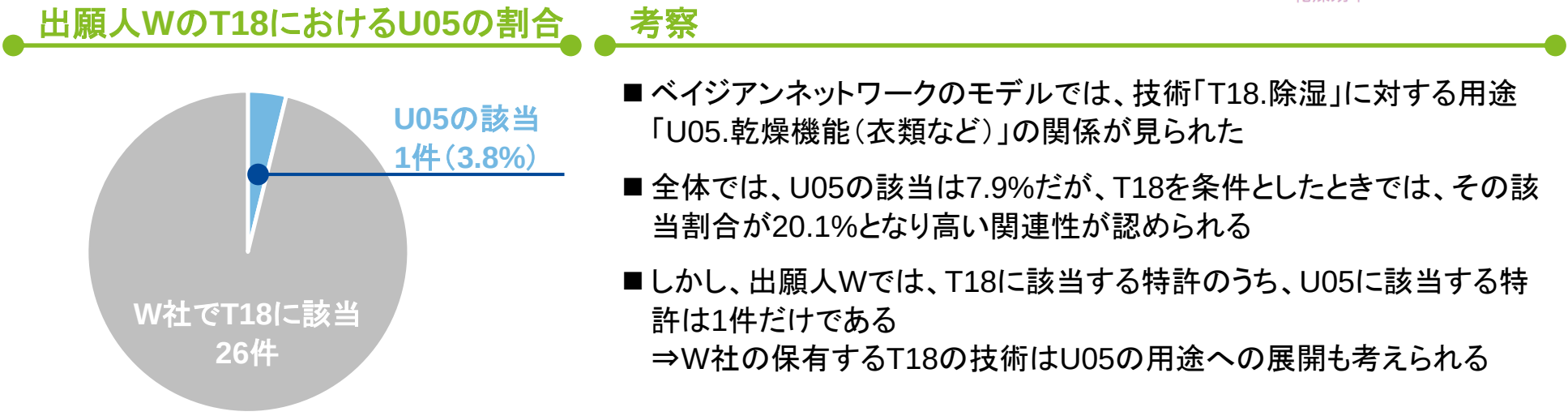
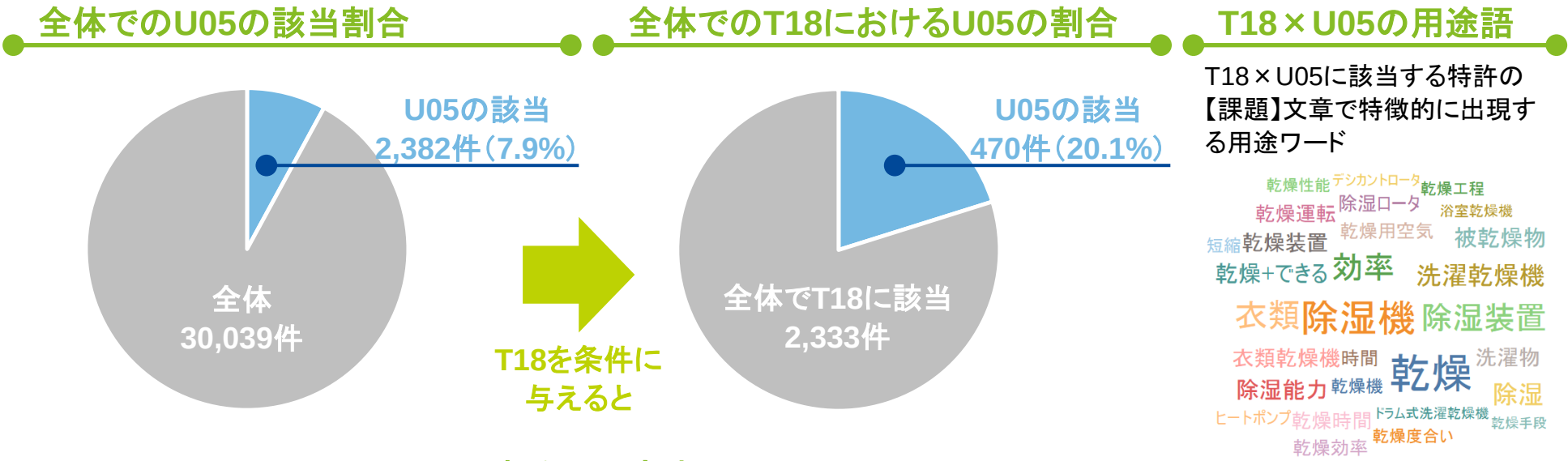
「T32.塵埃分離」と関係のある用途



「T35.運転と停止の制御」と関係のある用途



「T18.除湿」の技術の応用先として「U05.乾燥機能」の用途は高い関連性がありますが、出願人Wの保有するT18ではそれがほとんどなく、新規用途となる可能性があります



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乾燥処理技術は、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

T18がU05で応用されている例

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。

出願人Wの保有するT18の例

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】

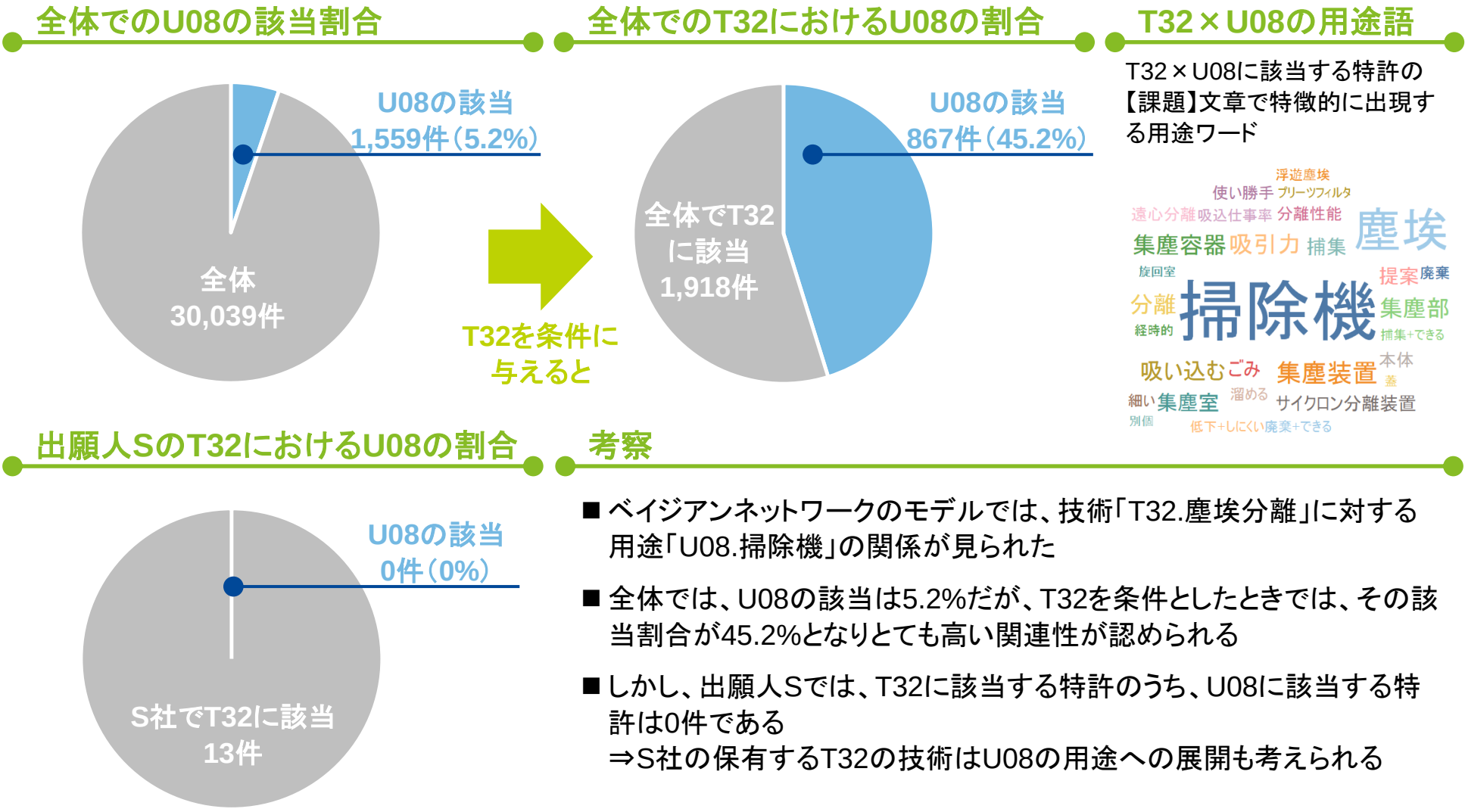
処理液の厚みムラを低減するとともに処理液による用紙のコックリングを低減することで、高品質かつ高速の画像記録を可能とするインクジェット記録装置及び画像記録方法を提供する。

【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

「T32.塵埃分離」の技術の応用先として「U08.掃除機」の用途はとても高い関連性がありますが、出願人Sの保有するT32ではそれが全くなく、新規用途となる可能性があります



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を向上させることにも応用できるかもしれません

T32がU08で応用されている例

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】

塵埃を含む空気を旋回させ塵埃分離する略円筒状の1次旋回室と、1次旋回室に連通した2次旋回室と、1次旋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次旋回室や1次旋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。

出願人Sの保有するT32の例

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

Nomolyticsによる特許文書分析のまとめ

Nomolytics®: Narrative Orchestration Modeling Analytics

PLSA 確率の潜在意味解析

ベイジアンネットワーク

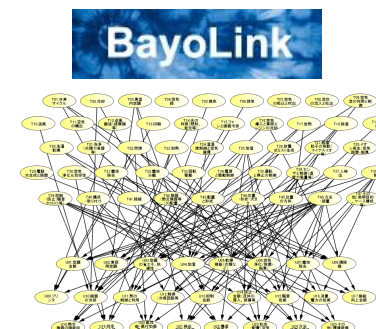
単語が出現する文脈を学習し、膨大な単語を複数のトピックにまとめる

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

トピック抽出

モデリング

Visual Mining Studio



テキスト情報内に潜
む複雑な要因関係を
構造化できる

条件を変化させたときの結果の挙動をシミュレーションできる

ある事象の発生確率
をコントロールする条
件を発見できる

テキストマイニングに加え、トピックを抽出するPLSAと、そのトピックの関係をモデル化するベイジアンネットワークを適用することで、特許情報に潜む特徴を分かりやすく把握できました

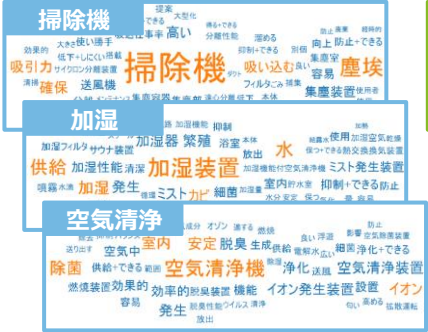
トピックの抽出

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

PLSAの適用

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
...	...	...

【課題】の要約文から用途トピックを抽出



【解決手段】の要約文から技術トピックを抽出



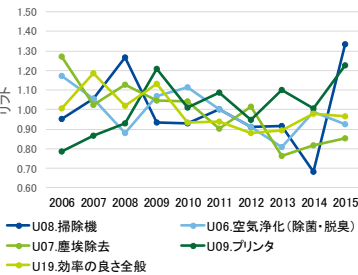
トピックのスコアリング

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	...	...	...	...	...	...	...	...

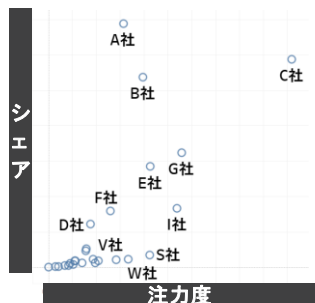
トレンドの分析

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



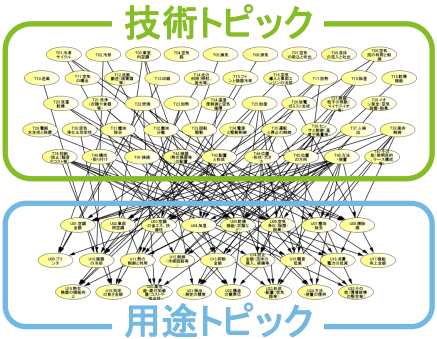
競合他社の分析

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



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出願動向を把握し、**開発戦略**や**提携戦略**を検討する

技術⇒用途 の分析

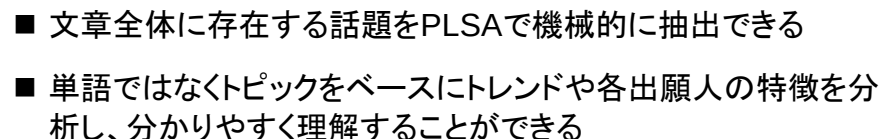
自社技術と関係する用途のうち未想定用途を発見し、**技術の新規用途展開**のアイデアを創出する

単語の共起ネットワーク



- 単語ベースのアウトプットから、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て、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

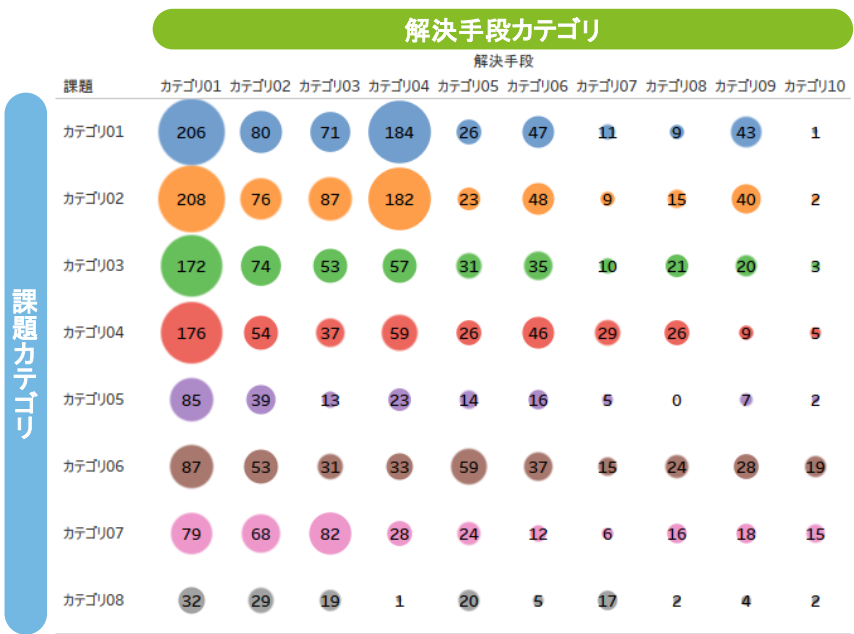
カテゴリ別のトレンド



用途と技術の統計的な関係を把握することで、各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

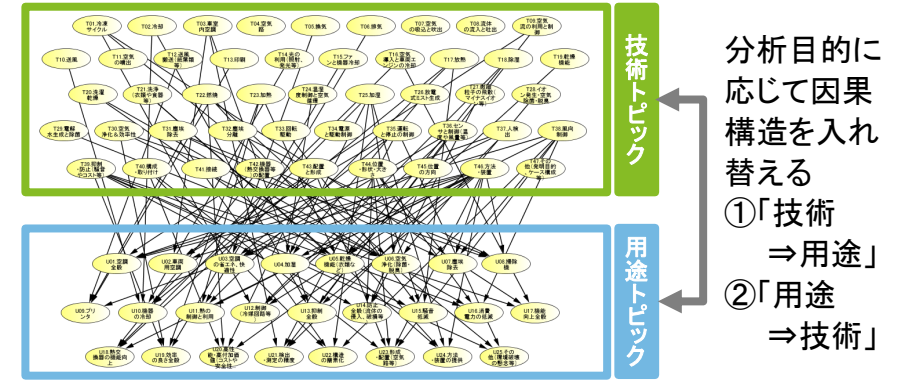
課題と解決手段のカテゴリ間のクロス集計



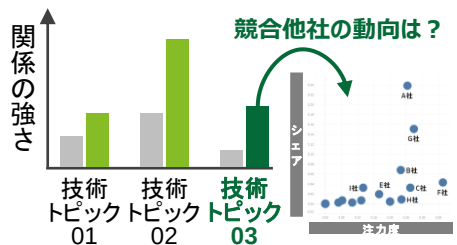
- 【課題】と【解決手段】それぞれに対して人がグルーピングして作成したカテゴリのクロス集計表を作成し、その対応関係を考察する
- その組み合わせで出願件数が多いからといって、統計的に意味のある関係であるとは限らない(全体的に出願件数が多い可能性もある)

Nomolyticsを適用した特許分析

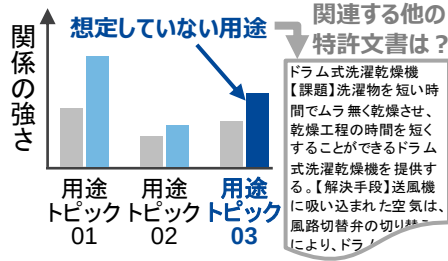
課題と解決手段のトピック間の統計的な因果関係モデル



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、各技術のトレンドや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

新技術

PCSA(確率的因果意味解析)

Nomolyticsは全体を表すトピックを抽出してから属性との関係进行分析しましたが、PCSAはその属性の特徴の要因となるトピックを最初から抽出し、より効果的な特徴の探索を行います

Nomolyticsは

データ全体を表現するトピックを抽出

してから

属性との関係进行分析

すると

属性の特徴を把握

できた

PCSAは

属性の特徴を表現する

ような

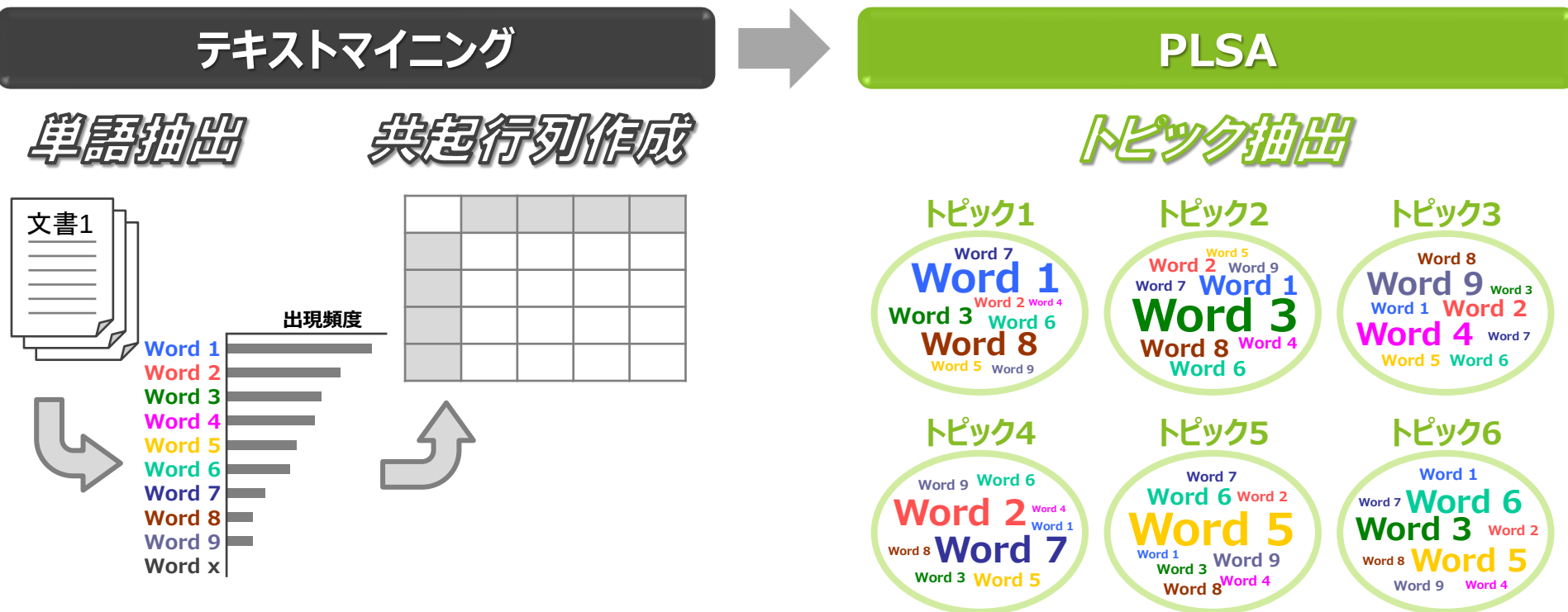
偏ったトピックを抽出

してから

より顕著な属性との関係进行分析

する

PLSAでは、テキストマイニングで抽出された単語で構成された共起行列をインプットにして、単語をトピックに集約し、テキストデータの全体像をシンプルに把握します



テキストデータにテキストマイニングを実行して単語を抽出し、その単語の共起頻度を集計した共起行列を作成する

作成した共起行列にPLSAを適用し、単語をトピックに集約する(使われ方の似ている単語をその重みと共にまとめる)

ターゲット事象の該当データと非該当データからそれぞれ構築した2つの共起行列の差分にPLSAを適用することで、そのターゲット事象に影響を与えるトピックを優先して抽出します

PCSA[®] (Probabilistic Causal Semantic Analysis : 確率的因果意味解析)

テキストマイニング

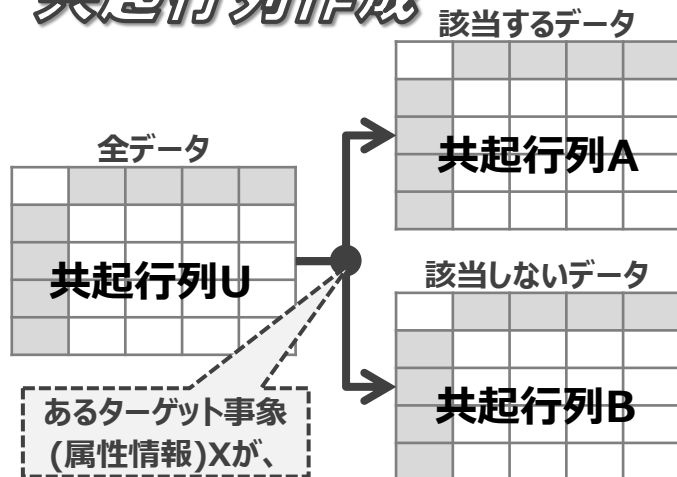
PLSA

単語抽出



全データから構築した共起行列Uを、あるターゲット事象(属性情報)Xが該当するデータから構築した共起行列Aと、該当しないデータから構築した共起行列Bに分割し、その2つの共起行列の差分を取った共起行列(A-B)に対してPLSAを適用する

共起行列作成



PLSAの適用

トピック抽出



ターゲット事象Xの該当有無に影響を与える潜在トピックを優先的にテキスト情報から抽出できる

ターゲット事象 X

【人工知能学会 2018年度全国大会優秀賞 受賞】

Nomolyticsでは教師なし学習(PLSA)を終了してから、その結果を使った教師あり学習を実行していますが、PCSAでは教師の情報を考慮したような教師なし学習を実行しています

Nomolytics

テキスト
マイニング

教師なし学習と教師あり
学習をそれぞれ独立させ
て順番に実行している

PLSA
確率的潜在
意味解析

教師なし学習

ベイジアン
ネットワーク

教師あり学習

PCSA

テキスト
マイニング

教師あり情報を差し込んだ
教師なし学習を実行し
ているようなイメージ

差分
共起行列

教師あり

PLSA
確率的潜在
意味解析

教師なし学習

PCSAを適用した特許分析事例

「風」「空気」に関する特許文書データの分析

「風」「空気」に関する10年分の特許データ30,039件を対象に、ターゲット事象を「出願年が2013年以後」に設定し、近年上昇傾向あるいは下降傾向にある技術トピックを抽出します

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～2015年

■ 抽出方法

- PatentSQUAREを使用

■ 抽出結果

- 30,039件



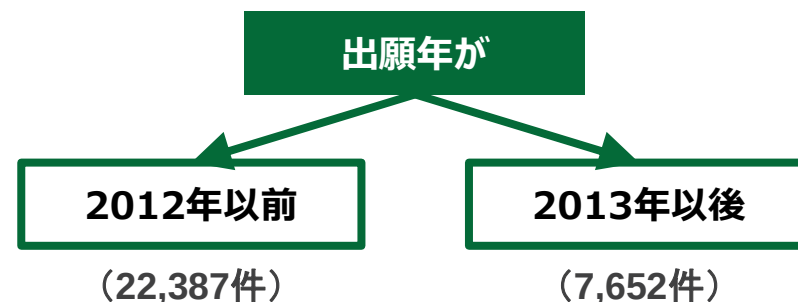
分析対象

■ 要約文の【解決手段】に記載されている文章

※上記書式で記載されていないものは要約文をそのまま使用

【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。

ターゲット事象



①2012年以前データ、②2013年以後データで作成した2つの共起行列の差分にPLSAを適用することで、2013年前後の出願に傾向のある技術トピックを優先的に抽出します



①出願年が**2012年以前**のデータ
22,387件(文章数:33,283件)

単語	係り受け表現				
	空気↓吸い込む	吸い込む↓空気	連↓通す	備える↓構成	⋮
	配置	144	139	109	102
	供給	95	65	53	81
	内部	93	64	92	44
送風機	145	134	79	76	
...					

②出願年が**2013年以後**のデータ
7,652件(文章数:11,831件)

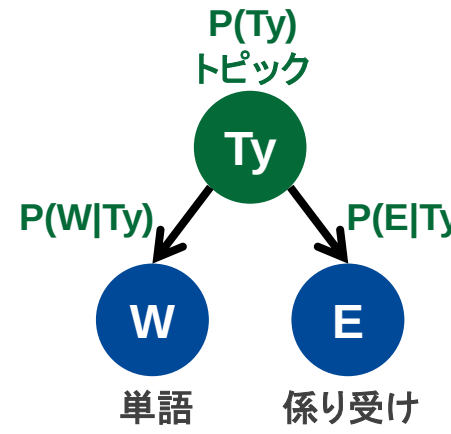
単語	係り受け表現				
	空気↓吸い込む	吸い込む↓空気	連↓通す	備える↓構成	⋮
	配置	66	52	41	43
	供給	32	36	14	27
	内部	41	43	34	20
送風機	80	63	33	19	
...					

①2012年までの共起行列と
②2013年からの共起行列の
差の絶対値を計算した共起行列
を作成する

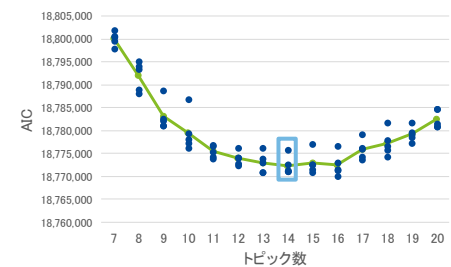
単語	係り受け表現				
	空気↓吸い込む	吸い込む↓空気	連↓通す	備える↓構成	⋮
	配置	14.8	2.6	2.3	6.7
	供給	1.8	12.9	4.8	1.8
	内部	7.9	20.3	1.3	4.4
送風機	28.5	15.4	4.9	8.0	
...					

ただし、2つの共起行列は異なる
文章数のデータから作成されて
いるので、①2012年までの共起
行列の頻度を2つの文章数の比
率(11,831/33,283)で重み調整
してから、②2013年からの共起
行列との差を計算する

差分の共起行列に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



各トピックについて以下の3つの
確率が計算される

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

トピックにおける $P(W|Ty)$ と $P(E|Ty)$
からトピックの意味を解釈する

トピックTy10
 $P(Ty10) = 6.3\%$

確率	単語	確率	係り受け
3.4%	塵埃	0.9%	付着-塵埃
1.8%	送風機	0.8%	塵埃-除去
1.7%	掃除機	0.8%	吸い込む-塵埃
1.6%	吸い込む	0.7%	塵埃-含む
1.3%	分離	0.7%	塵埃-吸い込む
1.3%	フィルタ	0.7%	塵埃-分離
1.3%	捕集	0.6%	発生-送風機
1.0%	集塵部	0.6%	含む-空気
...	...	...	...

確率の高い構成要素から、Ty10
のトピックは「塵埃の分離」に関す
る技術トピックと解釈できる

冷凍サイクルや冷却ファン、衣類乾燥、燃焼、イオン発生、電解水の生成、塵埃の分離、羽の回転、検出と制御、車内空調の配置などの技術が14個抽出されました

Ty01.冷凍サイクル

温度圧縮 送風 吐出熱 膨張弁 接続
切り替える 車室内 冷媒回路 冷却 放熱器 蒸发器
熱交換器 室内熱交換器 圧縮機 減圧 室外熱交換器
熱交換 吸い込む 冷凍サイクル 冷媒
停止 流れる 凝縮器 循環制御 放熱
加熱 通過

Ty02.空気の冷却

通す 室内 屋外熱交換 外部
導入 配設 内部冷却 外気 送風機 空気風路
排気風路 吸い込む 排気口 吸込口
通過 排出 取り込む 冷却装置 ケース排気
吹き出す 吹出口 循環設置 開口
供給 連通

Ty03.冷却ファン

光源 加熱 載置
熱 吸い込む 取り込む 光 収納 下方 ファン
配置 供給 冷却ファン 筐体 吸気口
冷却 電気部品 外部上方 ヒートシンク
冷却風 内部送風位置 収容
覆う 排出

Ty04.空気流 (吸込と吹出)

風向ける 室内 熱交換 送風 送風機 回転
下方取り付ける 空気流 送風ファン 室内空気 筐体内 吸込口
熱交換器 送風ファン 室内空気 筐体内 吸込口
吸い込む 吹出口 空気調和機
上下風向板 室内機 吹き出す 空気吸込口
空気風路 風向板
ファン 方向 筐体 位置 配置

Ty05.紙葉類の搬送

投入下流 開く 移動
到達用紙 センサ送り込む 定着部 制御部
加熱送風部 定着装置 トナー像 シート
送風管 画像形成装置 検出 搬送
空気流 通過 上流 吹き付ける 制御
定着開閉弁 紙葉類生ずる 閉じる 接続部

Ty06.衣類乾燥

内部制御
配設 排気口 送風ファン 水槽 収容 ヒートシンク
排出加熱手段 衣類 ヒートポンプ装置 放熱器
回転ドラム 乾燥用空気 送風 循環風路
乾燥運転洗濯乾燥機 乾燥室 送風手段
衣類乾燥機 洗濯物 加熱外槽 循環除湿
接続ドラム 乾燥

Ty07.空気の燃焼

水 排ガス 向上 熱風 温度 生成
外部発生 使用 提供 圧縮空気 送風機
燃焼 排出燃焼用空気 防止 供給
装置加熱 噴出 燃焼室 燃焼ガス
バーナ 混合 燃料放出熱 ノズル
乾燥

Ty08.イオン発生

付着 放電高電圧 電圧 生成 オゾン帯電
保持 吹出口 送風 空気中
電極 発生 対向電極 分解 送風路
空気風路 イオン発生装置 静電霧化装置
臭気成分 印加 供給 放電電極 吸込口
配置 イオン 放出 水分 空気清浄装置
ミスト

Ty09.電解水の生成

室内 貯留
貯水室 貯水部 下方 ケース 配置
接触 生成 加湿手段 供給 貯水 電解水
送風手段 送風ファン 加湿 加湿装置
水 給水タンク 加湿フィルタ 空気風路
吹出口 送風機 送風 通過 吸込口
水槽 上方 回転浸漬

Ty10.塵埃の分離

開口排出
上流 捕集 流入口 配置 通過
下方分離 含塵空気 送風機 集塵部
集塵容器 塵埃 除去 吸い込む
接続フィルタ 集塵装置 吸込口 掃除機
着脱自在 集塵室 吸引力 付着 旋回 流入
発生 連通

Ty11.羽の回転

連結ファン
空気流 向かう 中心 沿う 吸込口 周囲
一端ケーシング モータ 羽根 羽根車 固定
軸方向 対向 回転軸 ロータ 取り付け
回転 ファンケース 径方向 外周 外周側
配置 固定子 支持位置 回転+できる
内部

Ty12.検出と制御

測定動作 開始
記憶温度センサ 算出 決定 人 駆動 検出
制御手段 温度 運転 判定
空気調和機 基づく 制御部 制御
回転数 制御装置 設定 風量 停止 応ずる
所定 室内出力 湿度 状態 取得

Ty13.車両用空調 の配置

流れ 下方 後方
冷却風 車両 送風 導入 開口
吹出口 送風機 延びる 配置 吹き出す
車室内 車両用空調装置 側空気風路
導く 向ける 下流 ダクト 方向 流れる
沿う 向かう 位置 内部 上流 通過 上方

Ty14.配置と形成

挿入
平行 背面 突出 底部 覆う 固定端部
収納 配置 沿う 方向 傾斜 延びる
上方 位置 熱交換器 内部 取り付け
開口部 下方 下面 互い 仕切板
開口 対向側面 フィン 間隔
前面

Nomolyticsの時と同様に全特許データに対して各トピックの該当有無を計算し、トピックと出願年の関係を示すリフト値から、2013年前後に対する各トピックの特徴を分析します

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	出願年 グループ	出願人	技術トピック Ty01	技術トピック Ty02	...	技術トピック Ty14
1	特願2006-XXXX	2006	①2012年以前	A社	1	0		0
2	特願2009-XXXX	2009	①2012年以前	B社	0	0		0
3	特願2011-XXXX	2011	①2012年以前	C社	1	0		1
4	特願2013-XXXX	2013	②2013年以後	D社	0	1		1
5	特願2014-XXXX	2014	②2013年以後	A社	0	1		1
...	...	...	...	...	...	...		...
30039	特願2012-XXXX	2012	①2012年以前	Z社	1	0		1

各トピックと2つの出願年グループの関係を示すリフト値を計算

$$\text{リフト値① (2012年以前)} = \frac{P(\text{出願年①} \mid \text{トピックTy}=1)}{P(\text{出願年①})}$$

$$\text{リフト値② (2013年以後)} = \frac{P(\text{出願年②} \mid \text{トピックTy}=1)}{P(\text{出願年②})}$$

①から②の増減率を計算

出願年2013年前後での各トピックのリフト値の増減率を集計したところ、PCSAで抽出されたトピックは増減率の高いものと低いものに集中しています

PCSAで抽出したトピックのリフト値増減率

PCSAで抽出した14個のトピックは、増減率の高いものと低いものに集中している

リフト値の増減率	①2012年 以前の リフト値	②2013年 以後の リフト値	トピック
48.5%	0.89	1.32	Ty13.車両用空調の配置
30.3%	0.93	1.21	Ty14.配置と形成
23.9%	0.94	1.17	Ty04.空気流(吸込と吹出)
17.7%	0.96	1.13	Ty12.検出と制御
14.9%	0.96	1.11	Ty11.羽の回転
13.1%	0.97	1.09	Ty03.冷却ファン
11.6%	0.97	1.08	Ty02.空気の冷却
7.0%	0.98	1.05	Ty01.冷凍サイクル
-0.3%	1.00	1.00	Ty05.紙葉類の搬送
-5.9%	1.02	0.96	Ty07.空気の燃焼
-8.8%	1.02	0.93	Ty09.電解水の生成
-17.0%	1.05	0.87	Ty10.塵埃の分離
-24.2%	1.07	0.81	Ty06.衣類乾燥
-28.7%	1.08	0.77	Ty08.イオン発生

Nomolyticsで抽出したトピックのリフト値増減率

全体の共起行列からPLSAで抽出した47個のトピックは、増減率の高いものから低いものまでばらついている

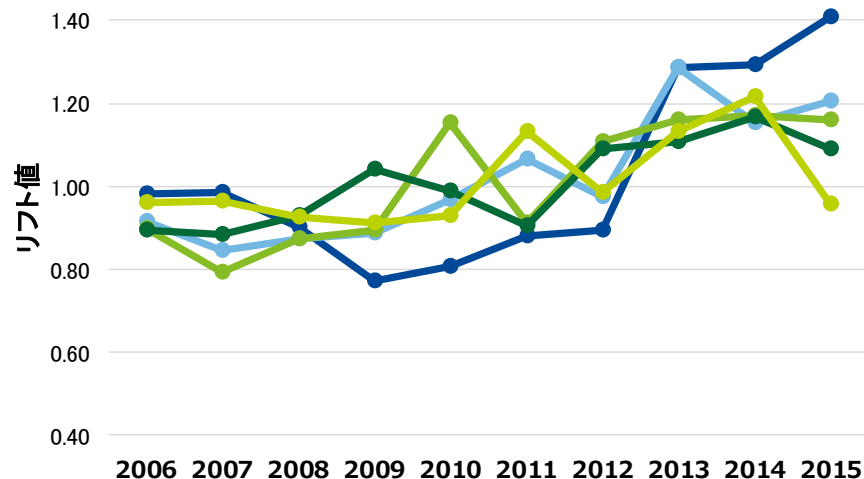
リフト値の増減率	①2012年 以前の リフト値	②2013年 以後の リフト値	トピック
19.3%	0.95	1.14	T44.位置・形状・大きさ
17.3%	0.96	1.12	T09.空気流の利用と制御
16.1%	0.96	1.12	T43.配置と形成
15.0%	0.96	1.11	T36.センサと制御(温度や風量等)
14.0%	0.97	1.10	T38.風向制御
13.5%	0.97	1.10	T16.空気導入と車両エンジンの冷却
...			
0.8%	1.00	1.01	T35.運転と停止の制御
0.4%	1.00	1.00	T42.機器(熱交換器等)の配置
0.0%	1.00	1.00	T06.排気
-1.1%	1.00	0.99	T46.方法・装置
-1.3%	1.00	0.99	T22.燃焼
-1.6%	1.00	0.99	T23.加熱
...			
-19.5%	1.05	0.85	T31.塵埃除去
-20.7%	1.06	0.84	T28.イオン発生・空気除菌・脱臭
-21.3%	1.06	0.83	T39.抑制・防止(騒音やコスト等)
-22.0%	1.06	0.83	T19.乾燥機能
-24.0%	1.07	0.81	T18.除湿
-27.4%	1.07	0.78	T26.放電式ミスト生成

※リフト値の増減率の高い順に並べている

トピックのトレンド

各出願年とのリフト値を集計して全体でのトレンドを見ても、上昇トレンド、下降トレンドを形成するトピックが抽出されていることが分かります

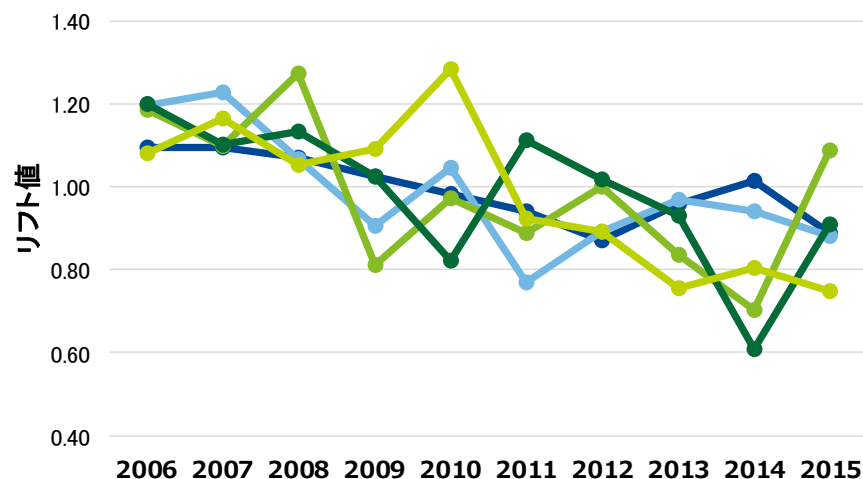
増減率上位5位のトピックのトレンド



2013年前後の増減率

Ty13.車両用空調の配置	48.5%
Ty14.配置と形成	30.3%
Ty04.空気流(吸込と吹出)	23.9%
Ty12.検出と制御	17.7%
Ty11.羽の回転	14.9%

増減率下位5位のトピックのトレンド



2013年前後の増減率

Ty07.空気の燃焼	-5.9%
Ty09.電解水の生成	-8.8%
Ty10.塵埃の分離	-17.0%
Ty06.衣類乾燥	-24.2%
Ty08.イオン発生	-28.7%

増減率ベスト1位は「車両用空調の配置」に関する技術、ワースト1位は「イオン発生」に関する技術で、それぞれのトピックに該当する特許の要約文の例を確認します

「Ty13.車両用空調の配置」の特許要約例

発明の名称

車両用空調装置

出願年

2015

【課題】

簡単な構成で、除湿暖房時における車両の燃費を向上させることができる車両用空調装置を提供する。

【解決手段】

冷却用熱交換器と、加熱用熱交換器と、冷却用熱交換器から加熱用熱交換器に至るヒータ通過経路と、加熱用熱交換器を迂回するヒータ迂回経路への空調風の流入割合を調整するエアミックスドアと、を備える車両用空調装置である。車両用空調装置は、加熱用熱交換器よりも空調風の流れの上流側で、かつ加熱用熱交換に面する位置に不動に設けられる除湿部材と、除湿部材の車両上方側を覆うように配置され、除湿部材の近傍から車両上方側への空調風の流れを制限する対流制限板と、を備える。

「Ty08.イオン発生」の特許要約例

発明の名称

扇風機

出願年

2007

【課題】

マイナスイオンを発生することができる扇風機において、寿命の長いイオンを生成する。

【解決手段】

放電電極に水分を供給する貯水タンクおよび吸水管とを備え、放電電極と対向電極との間に高電圧を印加することで放電電極に保持される水分を霧化させる静電霧化装置を具備する。したがって、人体に有益で、かつ消臭や除菌等に効果があるイオンが水に包まれることで、寿命の長いイオンミストとして空中に放出することができ、また水に包まれている分重いイオンミストを、扇風機によって広範囲に拡散することができる。

※対外説明用のため要約文は一部加工している

まとめ

AI技術を組み合わせて応用することで、膨大なテキストデータに潜む特徴や要因関係を可視化し、ビジネスアクションに有用なインサイトの獲得を図ります

Nomolytics®

テキストデータ全体を表すトピックを理解し、その平均的なトピックの特徴を様々な分析軸で探索できる

テキスト
マイニング

PLSA

ベイジアン
ネットワーク

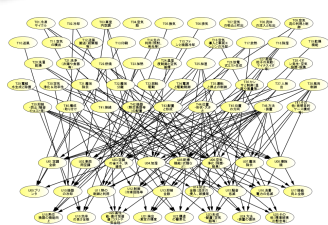
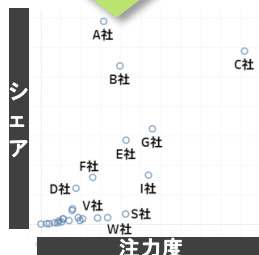
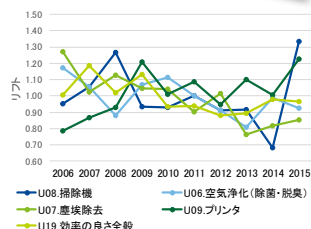
単語抽出

×

トピック抽出

×

モデリング

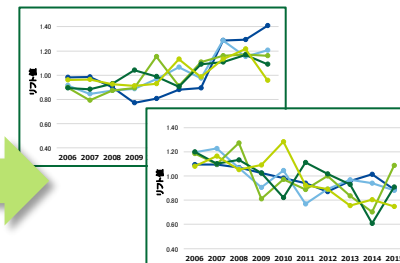
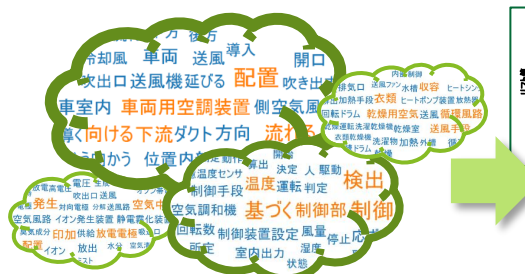
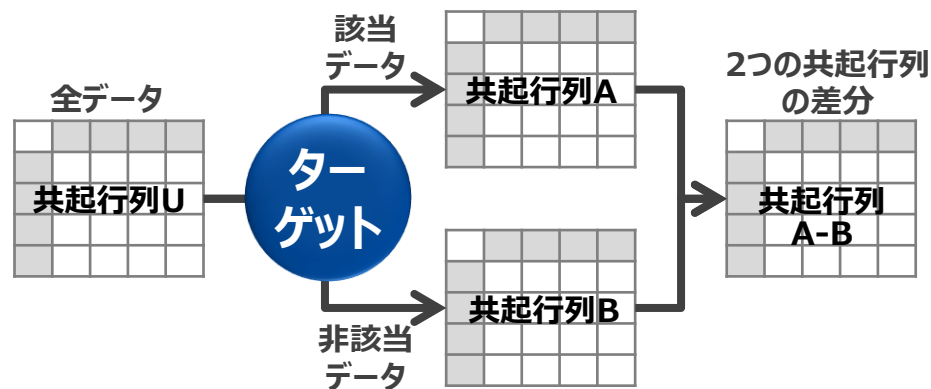


PCSA®

探索したい特徴に特化したトピックを優先的に抽出し、より顕著な要因を深く分析してインサイトを得る

テキスト
マイニング

PLSA



NomolyticsもPCSAも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

- 顧客の関心トピックのターゲット別把握
- 顧客目線での製品や競合の比較分析
- 満足度向上の要因の把握
- 価値観を理解したマーケティング検討



アンケート

- 自由記述の内容をトピック化
- 自由記述トピックを変数として扱うことで定型設問回答と一緒に分析可能
- 話題を生む要因の把握



コールセンター履歴

- 問い合わせ内容をトピック化
- 製品別・顧客別の問い合わせ特徴把握
- 問い合わせトピック等の条件から解約確率をシミュレーション



特許文書

- 課題と技術のトピックのトレンド把握
- 競合他社の技術動向把握
- 課題と技術のトピックの関係モデル化による保有技術の新規用途探索



営業日報

- 営業活動内容のトピック化
- 営業活動トピック等の条件から成約確率をシミュレーション
- 成約要因を把握した効果的な営業教育



有価証券報告書

- 各企業の事業内容をトピック化
- 事業トピックとそのトレンド把握
- 各種IR指標と事業トピックの関係分析
- 定性情報からの企業分析、業界分析



エントリーシート

- 志望動機やPR文のトピック抽出
- 記述内容からの学生の分類・振り分け
- 記述内容と入社後成果の関係分析
- 効率的な人材発掘



診療記録

- 診療記録、看護記録のトピック化
- 生活習慣と病状の関係分析
- 治療内容とその経過の関係分析
- 定性情報を用いた効果的な診療支援



問題発生レポート

- 不具合やヒヤリハット等のトピック抽出
- 作業環境等の条件から問題の発生確率をシミュレーション
- 効果的な製品や作業環境の改善支援

ビッグデータからインサイト獲得のためのダブルアプローチ

ビッグデータはそのままでは複雑で理解不能なので、まず理解できる形に抽象化して特徴を発見しますが、今度は抽象度が高くて業務に活用できないので、その特徴の個別のデータに着目して再度具体化します

特徴を発見しやすくするために抽象化する（量的分析に軍配を上げ平均的な存在の間にある普遍的な特徴を得る）

マクロ探索アプローチ

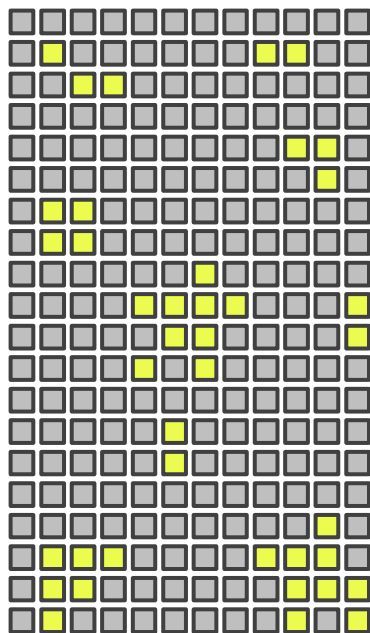
業務課題



確認したい
データの抽出

価値ある
個別データの
詳細調査と業務
活用への
昇華

膨大で複雑な ビッグデータ

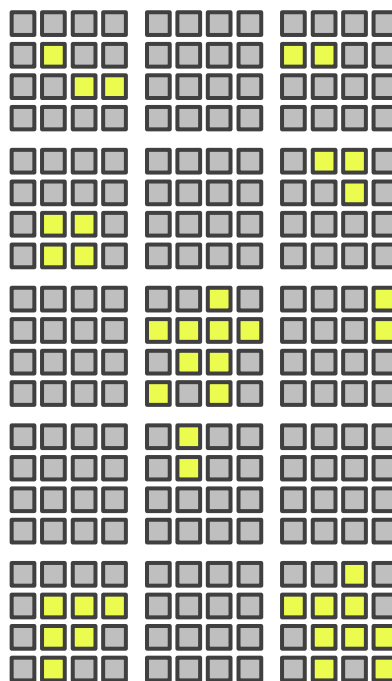


■ 価値あるデータ

特徴の似
ているデー
タの分類

特徴ピック
から個別
のデータの
絞り込み

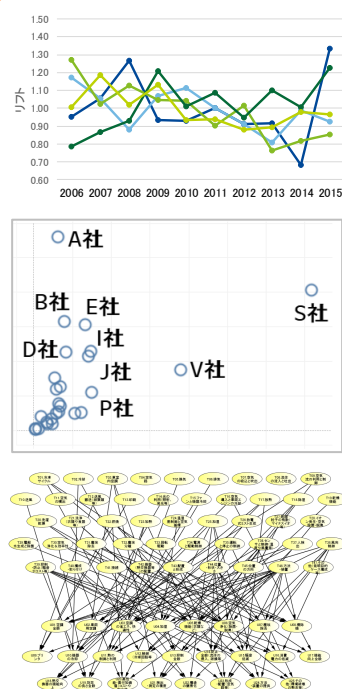
類型化



各トピック
の特徴の
可視化

注目すべ
きトピック
や要因条
件の発見

可視化・モデル化



ミクロ探索アプローチ

価値ある個別のデータを発見するために具体化する（質的分析に軍配を上げ平均の中の個別の特性を確認する）

昨年のユーザーコンファレンス2018では、PCSAを電気自動車関連の特許に適用した事例を発表しました。今年の人工知能学会ではさらに新しく開発したテキスト分析技術を発表します。

電気自動車関連の特許にPCSAを適用した事例

- 数理システムユーザーコンファレンス2018にて、電気自動車関連の特許文書データにPCSAを適用した分析事例を発表しました
- ホームページにて講演資料を公開しています



MSI
USER CONFERENCE
2018

数理システムユーザーコンファレンス
あらゆる分野のソリューションに対応。データ分析現場の最先端事例が知見を拡充。

参加無料：完全事前申込制
【今年より当日受付がなくなりました】

高岡 隆博：製造業における機械学習・IoTの適用とデータサイエンティスト育成 - JFEスチールの事例紹介 - JFEスチール株式会社 茂森 弘博 氏

会場：東京コンファレンスセンター・品川
日時：11.22 (THU) 10:15~17:00 (開場 9:30)

NTT DATA
〒160-0016 東京都新宿区西新宿35 住友不動産11F
Tel. (03) 3358-6681 Fax. (03) 3358-1727
E-mail: msc18@msi.co.jp 申込締切はこちら ▶ <http://www.msi.co.jp/>



さらに新しく開発したテキスト分析技術の発表

- 2019年度人工知能学会全国大会(第33回)にて、さらに新しく開発したテキスト分析技術を初めて発表します(特許出願準備中)
- 2019年度人工知能学会全国大会(第33回)
 - 開催期間：2019年6月4日(火)～6月7日(金)
 - 開催場所：新潟県新潟市(朱鷺メッセ 新潟コンベンションセンター)
 - <https://www.ai-gakkai.or.jp/jsai2019/>



2019年度 人工知能学会全国大会 (第33回)

2019年6月4日(火)～6月7日(金) 新潟県新潟市(朱鷺メッセ 新潟コンベンションセンター)

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。

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

株式会社アナリティクスデザインラボ

