



Analytics Design Lab

株式会社NTTデータ数理システム
ユーザーコンファレンス2019

ビッグデータ×AI時代のテキストリファインング技術

特許文書を例にしたテキストデータウェアハウスの獲得

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11月22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月に設立した企業様のデータ分析・活用を支援させて頂くコンサルティング会社です

株式会社アナリティクスデザインラボ

企業様のデータ分析・活用の支援をさせて頂く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、というデータ分析を活用するプロセスを企業様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様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
URL	http://www.analyticsdlab.co.jp/

代表取締役 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

➢ 2010年4月～2012年3月

独立行政法人日本学術振興会 特別研究員に採用

■ 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

弊社が分析を実施しご提供する「分析受託」、お客様が実施される分析を助言する「アドバイザリー」、弊社実施の分析をお客様にトランスファーする「テラー研修」をご用意しています

分析受託 サービス

お客様のデータをお預かりして
弊社がデータ分析を実施し、
結果をご報告します

- お客様の業務課題とご提供頂くデータに応じて、弊社がデータ分析の設計を行い、実行します
- 弊社による分析の実施結果をご報告し、その報告書を成果物としてご納品します
- 分析の実施にかかる期間(作業工数)から費用をお見積りします

アドバイザリー サービス

お客様ご自身で実施される
データ分析・活用のご助言、
ご指導をします

- お客様の業務課題の解決に効果的なデータ分析・活用についてご助言します
- お客様が実施される具体的なデータ分析の作業についてもご指導します
- 弊社がご納品する成果物はありません
- 1回〇時間の訪問助言を何回ご提供するのかわによって費用をお見積りします

テラー研修 サービス

弊社が実施した分析の内容を
お客様で実施できるように、
その手順を全てレクチャーします

- 「分析受託サービス」で弊社が実施した分析について、実施手順マニュアルや分析のプログラムファイルのご提供とともに解説し、お客様で同様の分析を実行できるように技術トランスファーします
- 「分析受託サービス」の費用に加え、マニュアルの作成や研修の実施などにかかる工数から費用をお見積りします

テキストリファインング

テキストマイニング × AIによるトピックベースの特徴可視化

テキストマイニングは単語をベースにテキストデータの全体像を把握しますが、大量の単語で可視化される結果はとても複雑で解釈が難しいため、単語をグルーピングすることが有効です

テキストマイニングのアウトプットの例

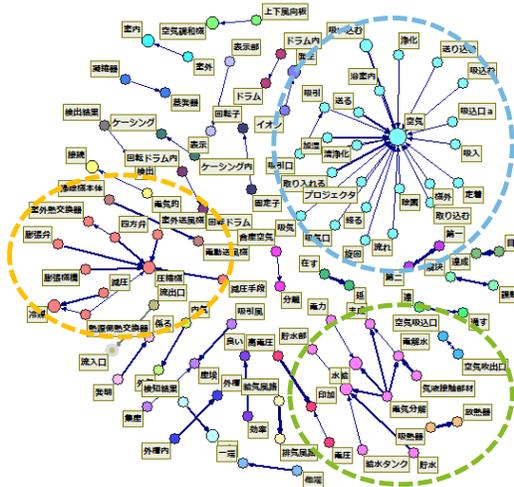
頻度集計

単語や係り受け表現の出現頻度を集計して、どのような記述が多いのか、おおまかな全体像を把握する



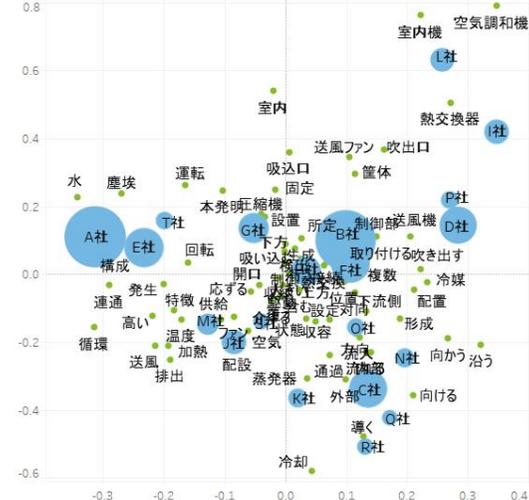
共起ネットワーク

同時に出現しやすい単語同士をネットワークでつなぎ、そのかたまりからどのような話題があるか考察する



コレスポネンス分析

属性情報と出現単語との対応関係を同じ平面上にマッピングし、その位置関係から属性の傾向を把握する



テキストマイニングの課題

結果が複雑で解釈がしづらい

- 単語をベースにした結果は複雑で、解釈が難しい
- 特に読み込むテキストデータがビッグデータになると、マイニングで抽出される単語も膨大となる

解決方法

単語をグルーピングする

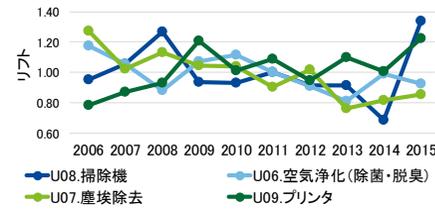
- 単語を人がグルーピングすることもあるが、その作業は主観的で作業負荷があまりにも大きい
- 機械的にグルーピングを実現することが望まれる

テキストリファイニングでは、テキストマイニングで抽出した大量の単語をトピックに集約し、そのトピックをベースに特徴を可視化することでテキストデータに潜む特徴をシンプルに理解します



室内における人の位置を確実に把握して人に直接冷風あるいは温風が当たらないように風の到達位置を制御することで快適空調を達成することができる空気調和機を提供する。室内機に設けられたファンの回転数と、室内機に設けられた上下羽根の角度を制御して、暖房時は人体検知センサにより検知された人の足元近傍に温風を到達させる一方、冷房時は人体検知センサにより検知された人の頭上上方に冷気を到達させるように設定した。

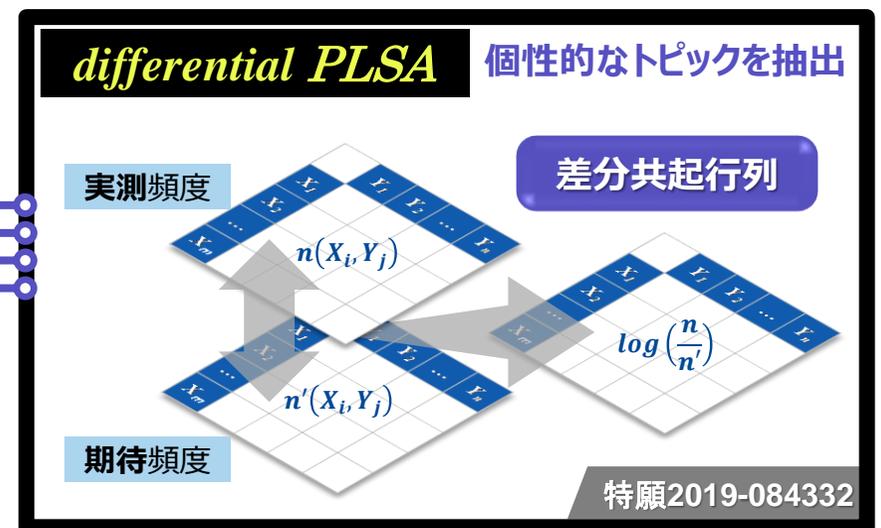
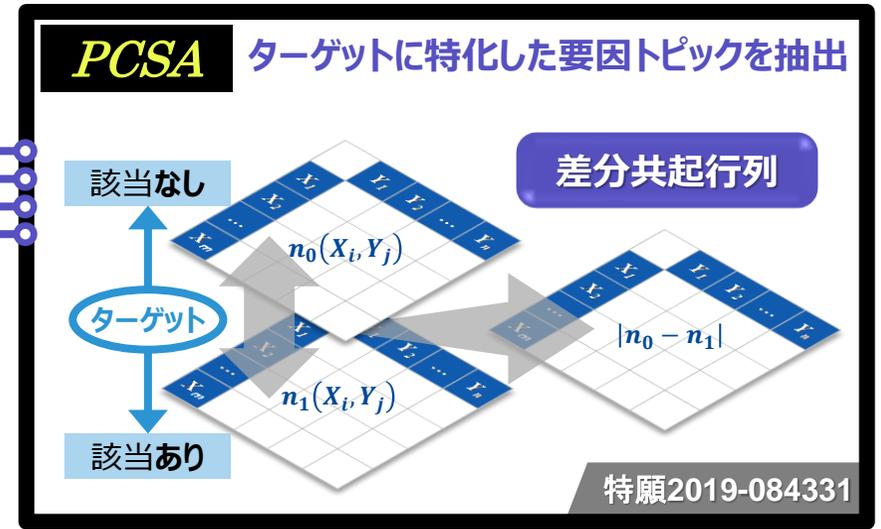
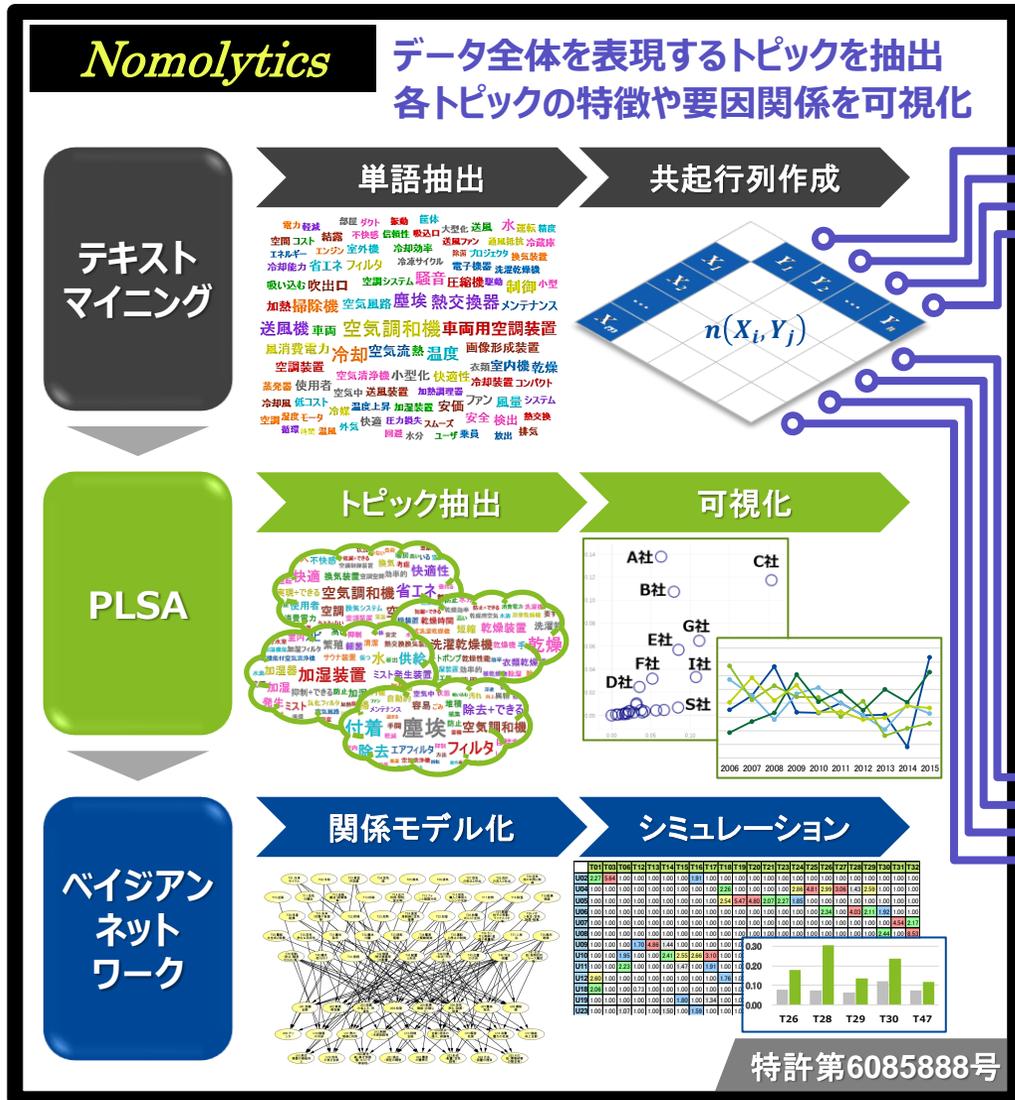
大量の抽出された単語のクラウド。中心には「空気調和機」があり、周囲には「熱交換器」、「送風機」、「加湿装置」、「除湿機」、「空気清浄機」などの関連用語が散らばっている。



テキストリファイニングについて開発した3つの技術

テキストリファイニングを実行する技術として、トピックモデルを中心とするAI技術を応用した

① Nomolytics、② PCSA、③ differential PLSAという3つの技術を開発しました



Nomolytics

テキストデータ全体を表すトピックを抽出してデータに潜む特徴を俯瞰する

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

単語抽出

Text Mining Studio



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

トピック化

Visual Mining Studio

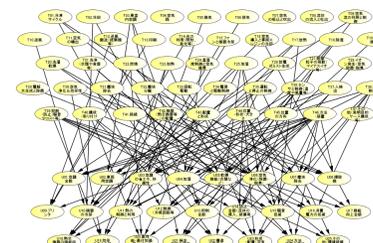


ベイジアンネットワーク

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

因果分析

BayoLink



膨大なテキストデータを人間が理解しやすい形に整理できる

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

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

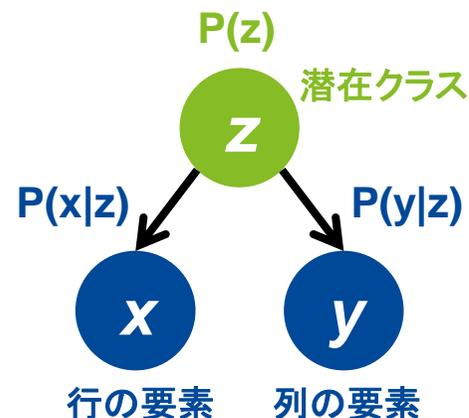
【共起行列】

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	1		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



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

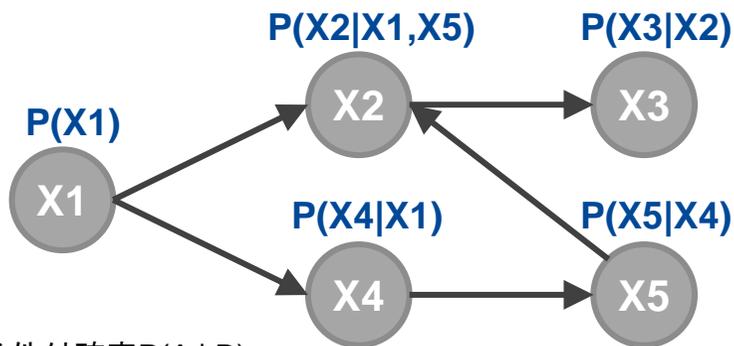
ソフトクラスタリングできる

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

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

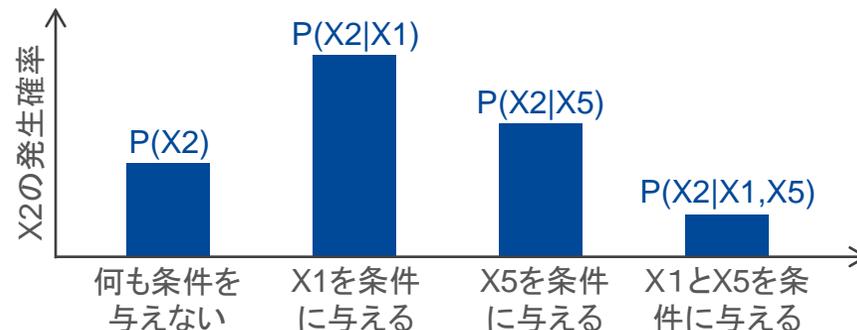
ベイジアンネットワークの概要

- 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、非線形の関係や相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

確率的因果関係と相互作用



- 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に対して相互作用がある(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相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

Nomolyticsを適用した特許分析事例

空調関連の特許データを用いた用途トピックと技術トピックの特徴分析

「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

【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。

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、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

トピックのスコアリング

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
...

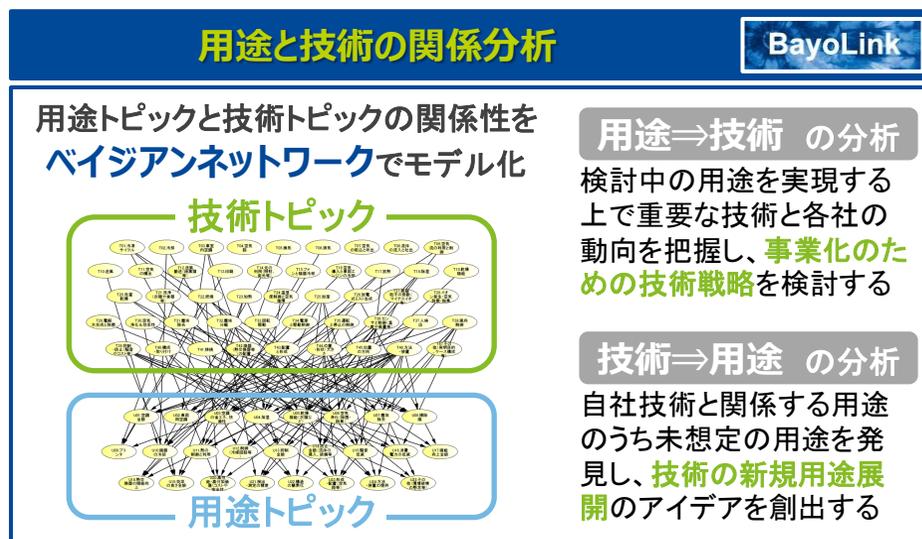
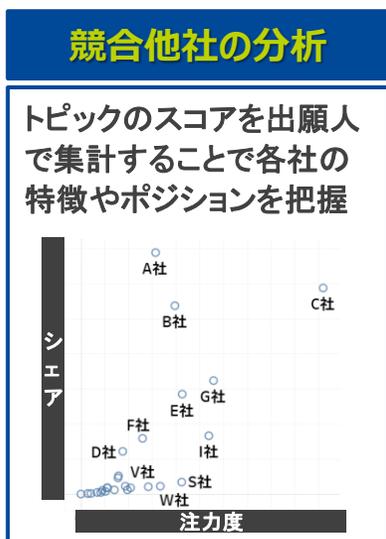
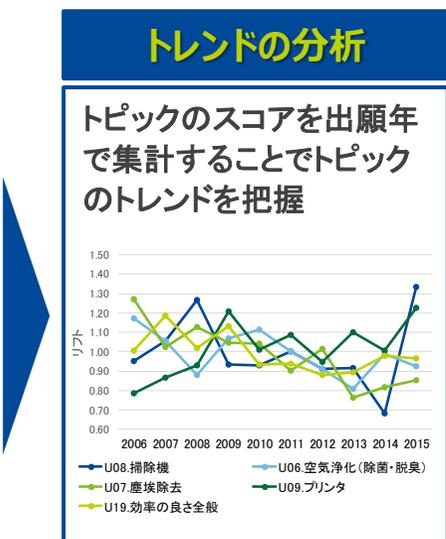
PLSAの適用

【課題】の要約文から用途トピックを抽出

【解決手段】の要約文から技術トピックを抽出

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



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

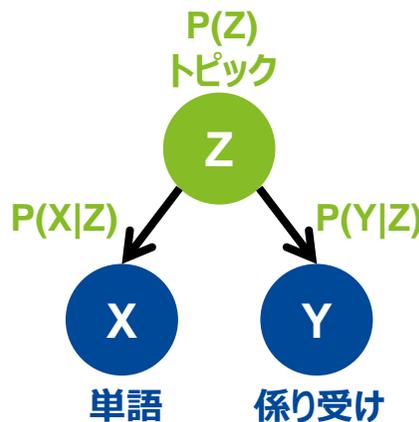
抽出した単語と係り受け表現に基づいて、「単語×係り受け」の共起行列(文章単位で同時に出現する頻度のクロス集計表)を作成する

	係り受け表現				
	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

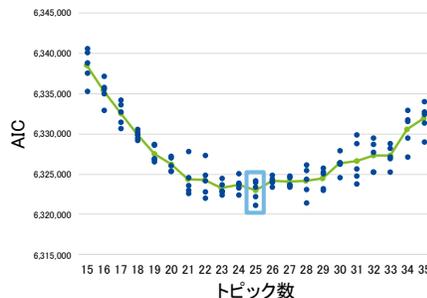
共起行列の構成(それぞれ頻度10件以上を対象)
 課題: 単語(3,256語) × 係り受け(2,084表現)
 解決手段: 単語(5,187語) × 係り受け(7,174表現)

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

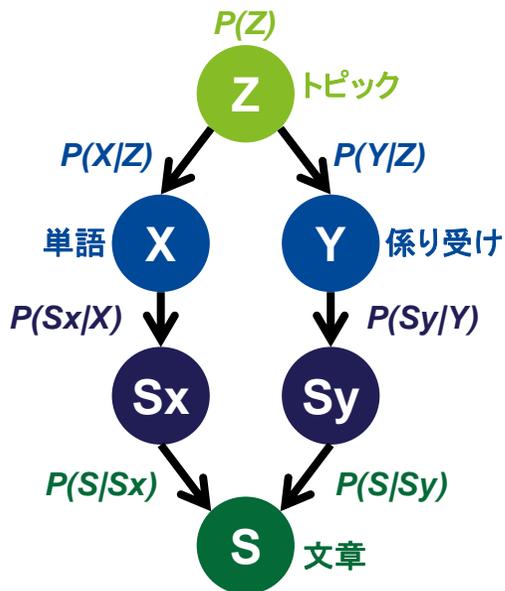
【解決手段】の文章からは、**空気の冷却や空気路、換気、放熱、除湿、乾燥、加湿、イオン生成、空気清浄、塵埃分離、センサと制御、構成や配置**などの技術が47個抽出されました

<h3>T01.冷凍サイクル</h3> <p>通過方向 膨張弁 蒸発器 凝縮器 冷媒回路 熱交換器 室内熱交換器 冷凍サイクル 蒸発器 圧縮機 冷媒 室外熱交換器 凝縮器 圧縮機 熱交換器 室内熱交換器</p>	<h3>T02.冷却</h3> <p>冷却器 冷却水 冷却装置 冷却装置 冷却水 冷却装置 冷却装置 冷却水 冷却装置 冷却装置 冷却水 冷却装置</p>	<h3>T03.車室内空調</h3> <p>車室内空調装置 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>	<h3>T04.空気路</h3> <p>空気路 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>	<h3>T05.換気</h3> <p>換気装置 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>
<h3>T06.排気</h3> <p>排気装置 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>	<h3>T07.空気の吸込と吹出</h3> <p>送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>	<h3>T08.流体の流入と吐出</h3> <p>送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>	<h3>T09.空気流の利用と制御</h3> <p>送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>	<h3>T10.送風</h3> <p>送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>
<h3>T11.空気の噴出</h3> <p>送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>	<h3>T12.送風搬送(紙葉類等)</h3> <p>送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>	<h3>T13.印刷</h3> <p>送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>	<h3>T14.光の利用(照射、発光等)</h3> <p>送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>	<h3>T15.ファンと機器冷却</h3> <p>送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>
<h3>T16.空気導入と車両エンジンの冷却</h3> <p>送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>	<h3>T17.放熱</h3> <p>送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>	<h3>T18.除湿</h3> <p>送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>	<h3>T19.乾燥機能</h3> <p>送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>	<h3>T20.洗濯乾燥</h3> <p>送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>
<h3>T21.洗淨(衣類や食器等)</h3> <p>送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>	<h3>T22.燃焼</h3> <p>送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>	<h3>T23.加熱</h3> <p>送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>	<h3>T24.温湿度制御と空気循環</h3> <p>送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>	<h3>T25.加湿</h3> <p>送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機 送風機</p>

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

- 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)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	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

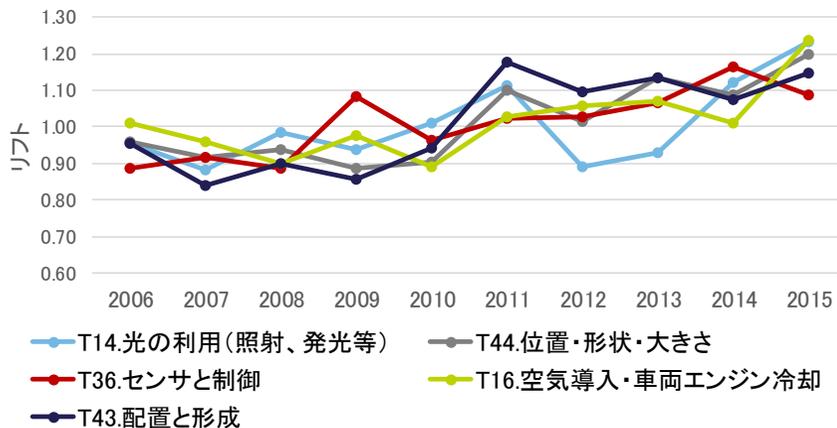
②出願人の分析

③用途と技術の関係分析

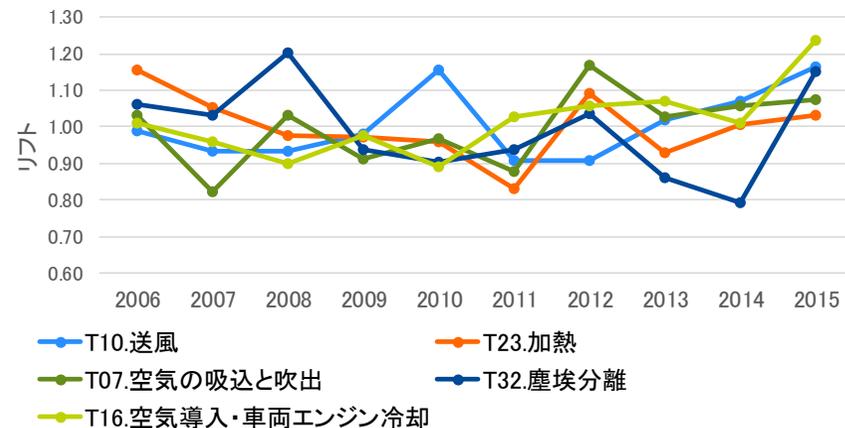
【①出願年の分析】 技術トピックの上昇トレンド

短期的に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トなどの光の利用に関する技術が上昇しています

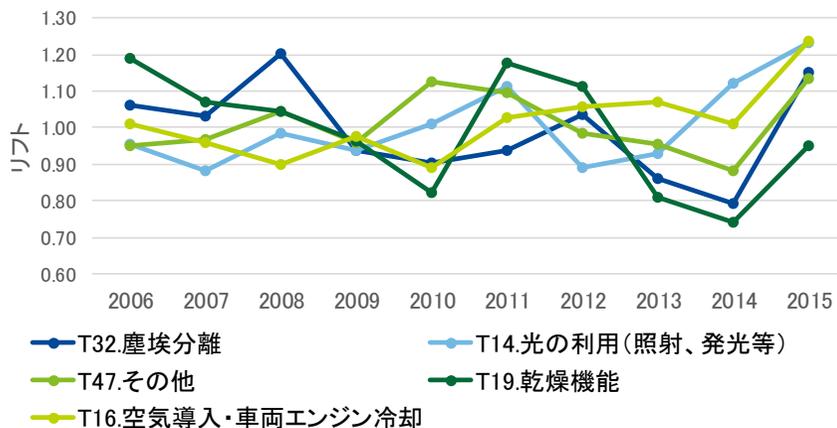
2006年からの上昇率 best5



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

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

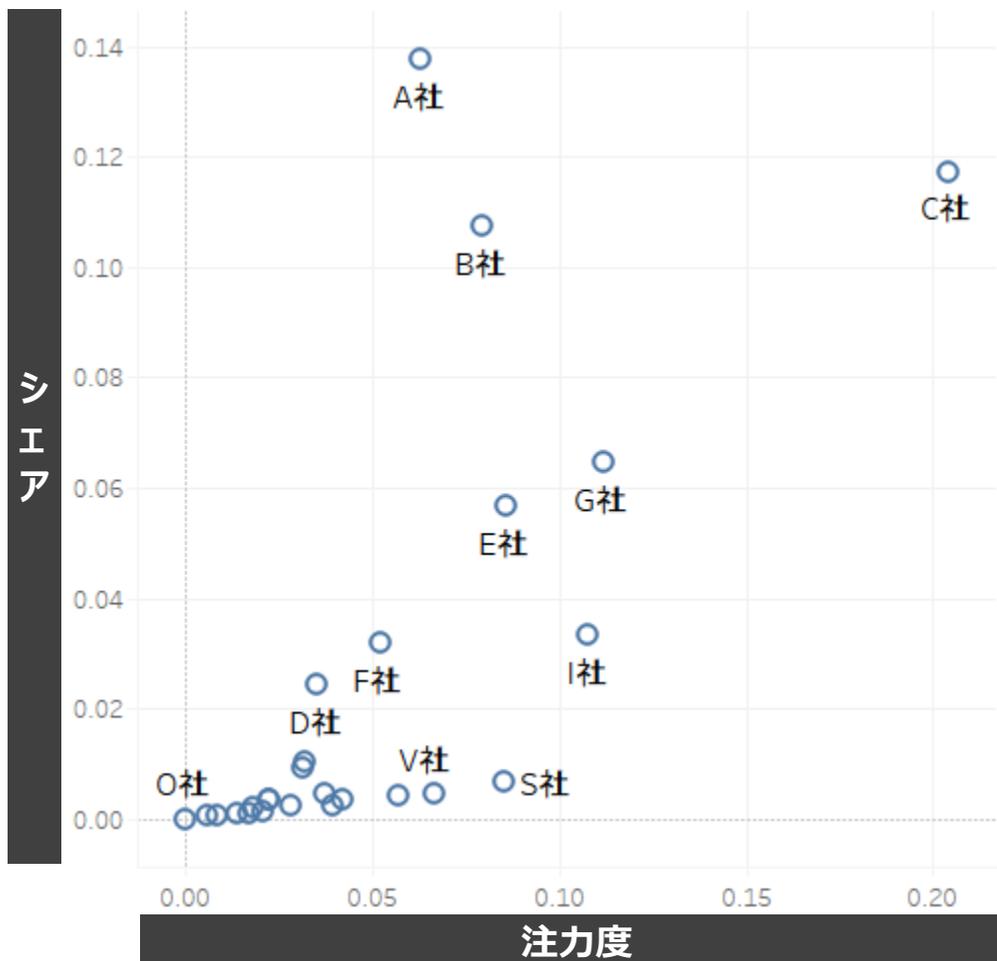
$$P(\text{出願年})$$

- 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

【②出願人の分析】 技術「T32.塵埃分離」の出願人のポジショニング

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

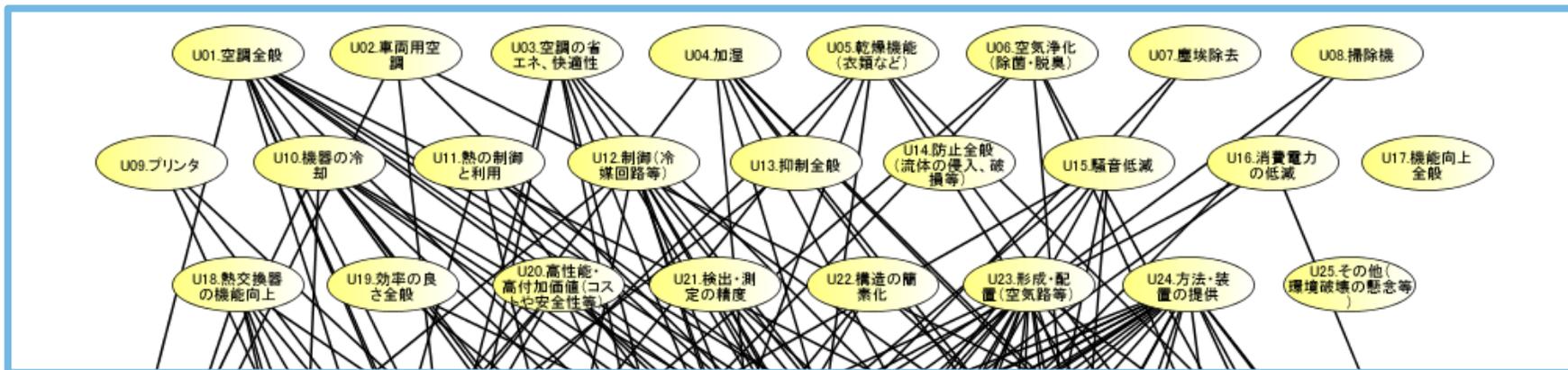
- C社は、高めのシェアを獲得しつつ、他社と比べて注力度がとても高く、高い技術力を保有していると考えられ、今後はよりシェアを伸ばすことで高シェア高注力度のポジションを確立することができると思われる
- A社とB社は、シェアは高いがC社に注力度でギャップがあるので、例えばE社、G社、I社などのように規模は中程度だが注力度は比較的高く、技術力があると思われる企業と連携することで、C社の上のポジションを狙うことができる可能性がある

注力度とシェア

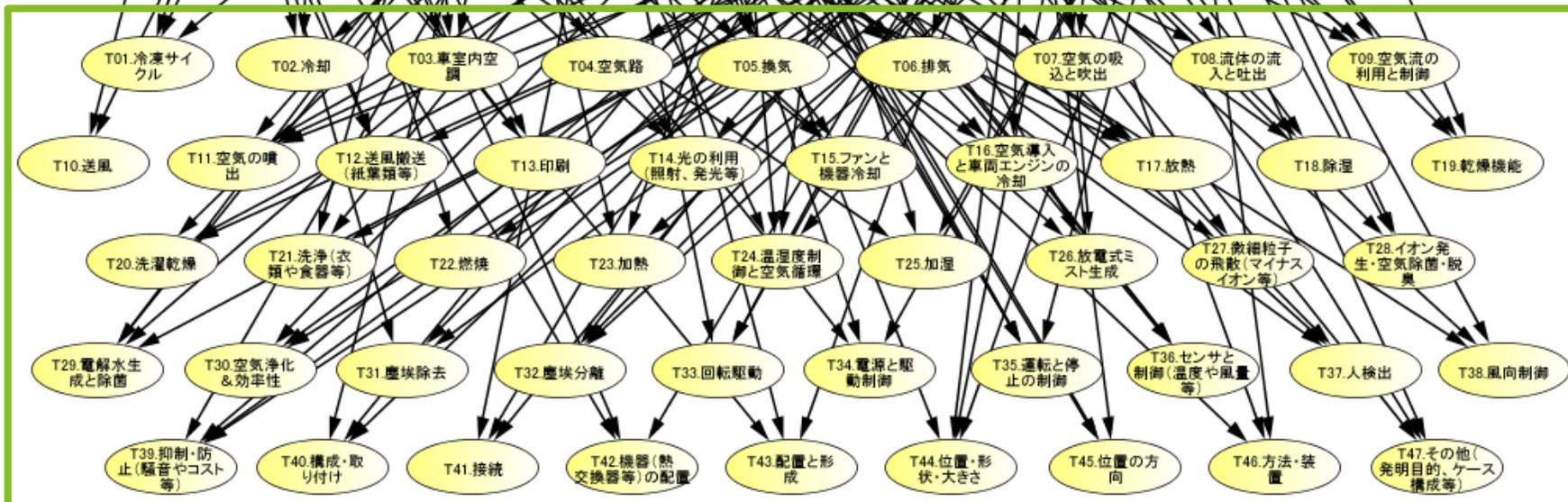
- **注力度**: $P(\text{トピック}T \mid \text{出願人}X)$
 - 出願人Xの出願特許の中で、どれくらいの割合がそのトピックTに該当するものか、つまり出願人がどれくらいそのトピックに注力しているのかを示す
- **シェア**: $P(\text{出願人}X \mid \text{トピック}T)$
 - トピックTが該当する特許の中で、どれくらいの割合がその出願人Xの出願によるものか、つまりトピックの中でどれくらいその出願人が占めているのかを示す

【③用途と技術の関係分析】 用途⇒技術の関係モデル

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

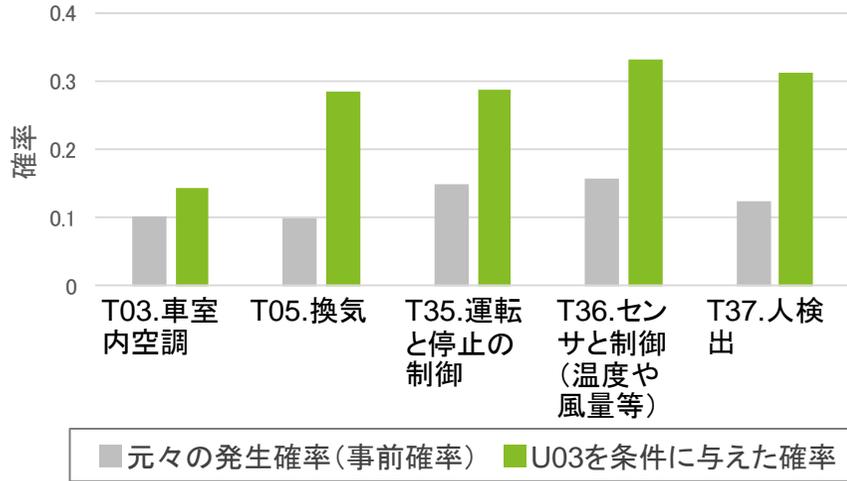


技術トピック

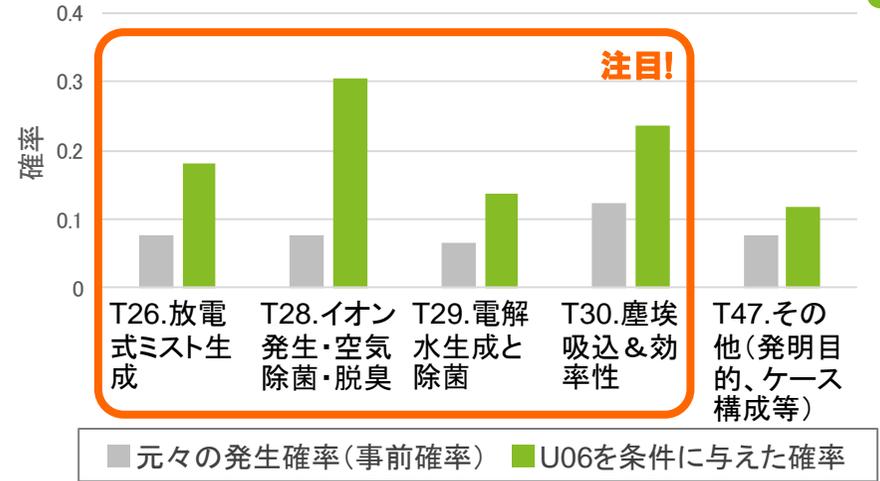
【③用途と技術の関係分析】 用途と関係のある技術の確認

1つの用途トピックを条件に与えたときに、ベイジアンネットワークで見られた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

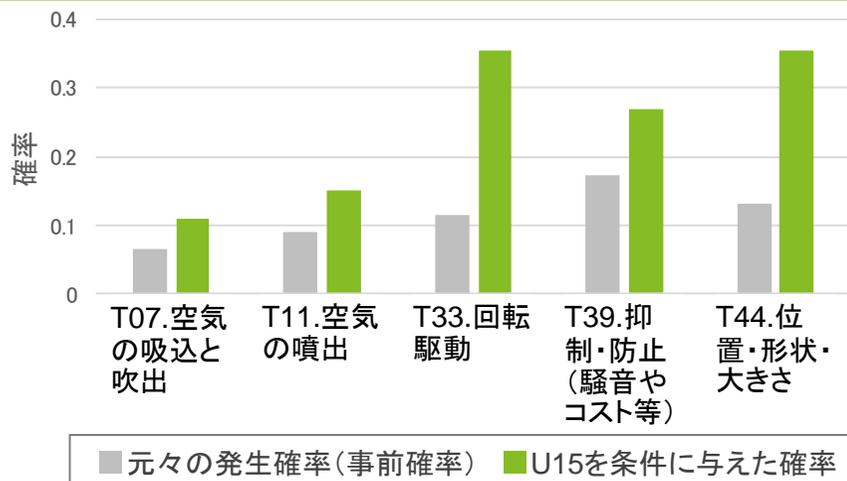
「U03.空調の省エネ、快適性」と関係のある技術



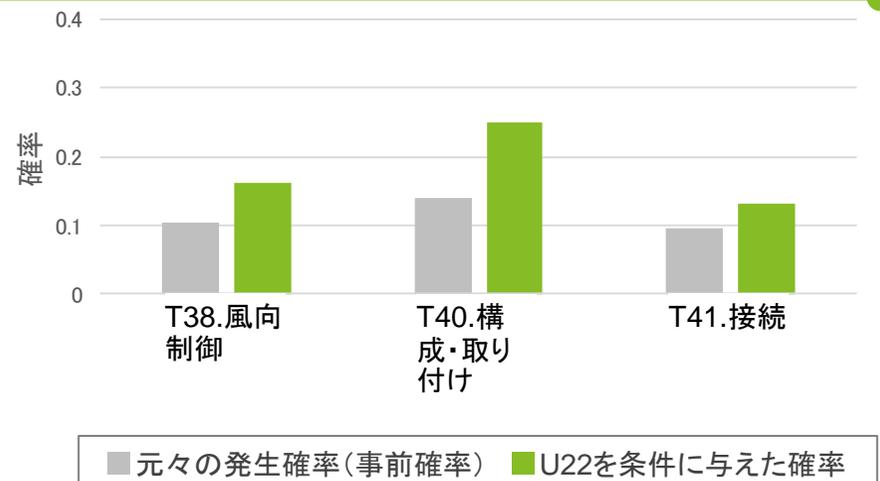
「U06.空気浄化(除菌・脱臭)」と関係のある技術



「U15.騒音低減」と関係のある技術



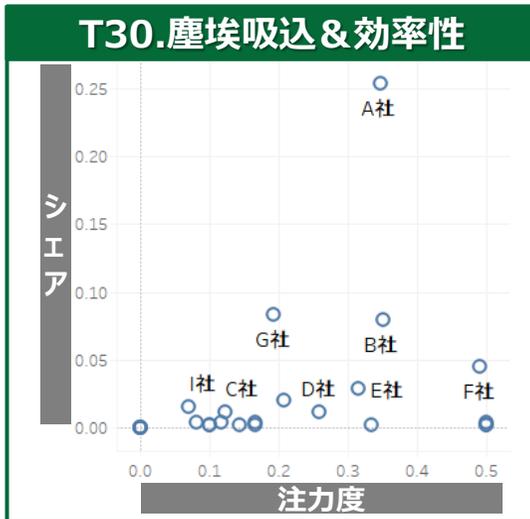
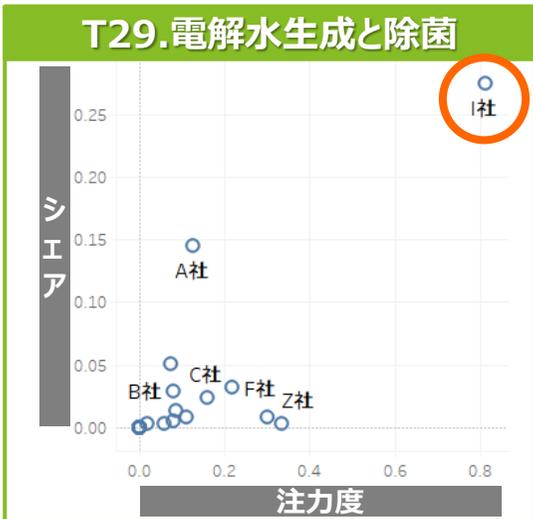
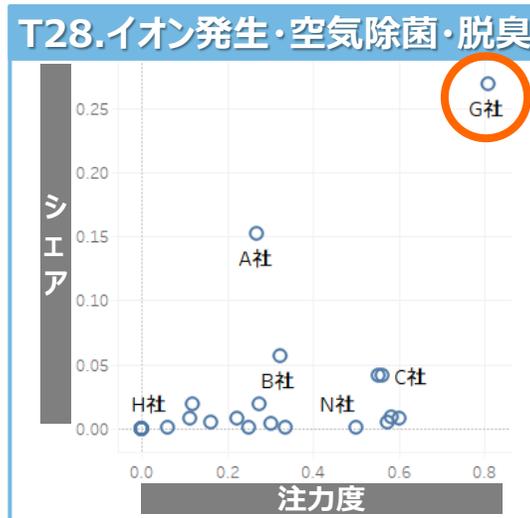
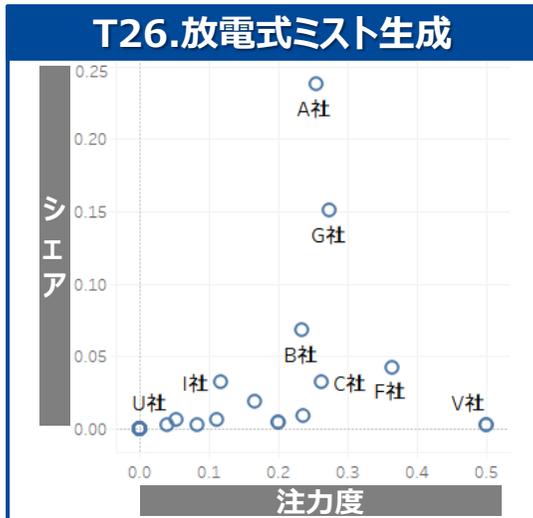
「U22.構造の簡素化」と関係のある技術



【③用途と技術の関係分析】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

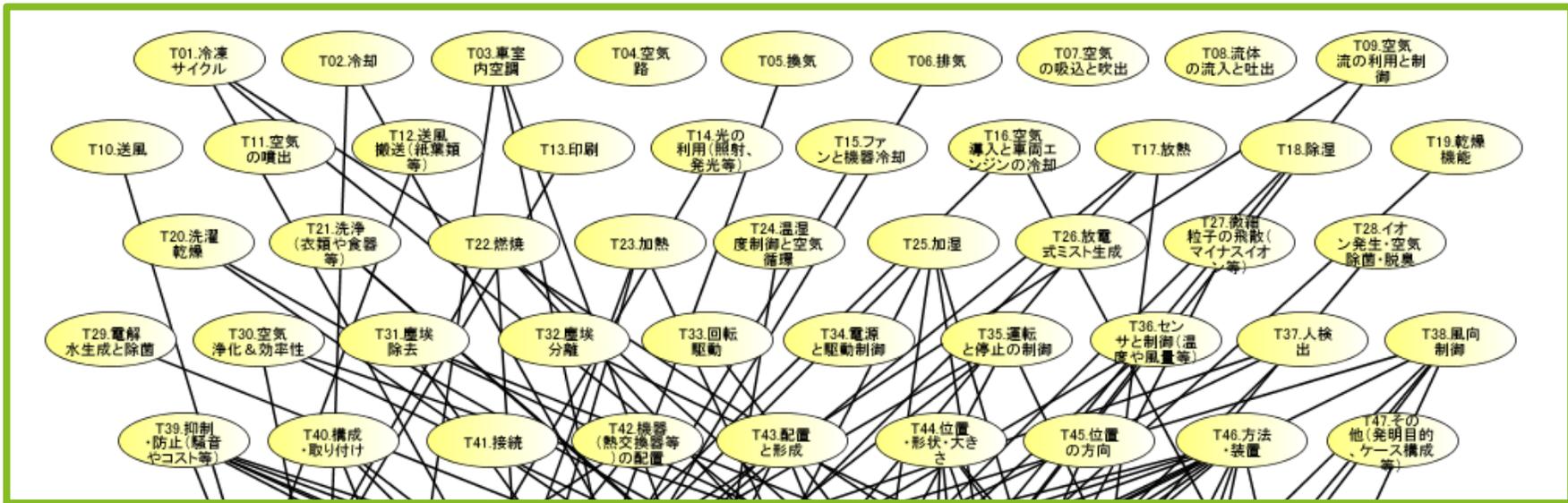


考察と戦略の検討

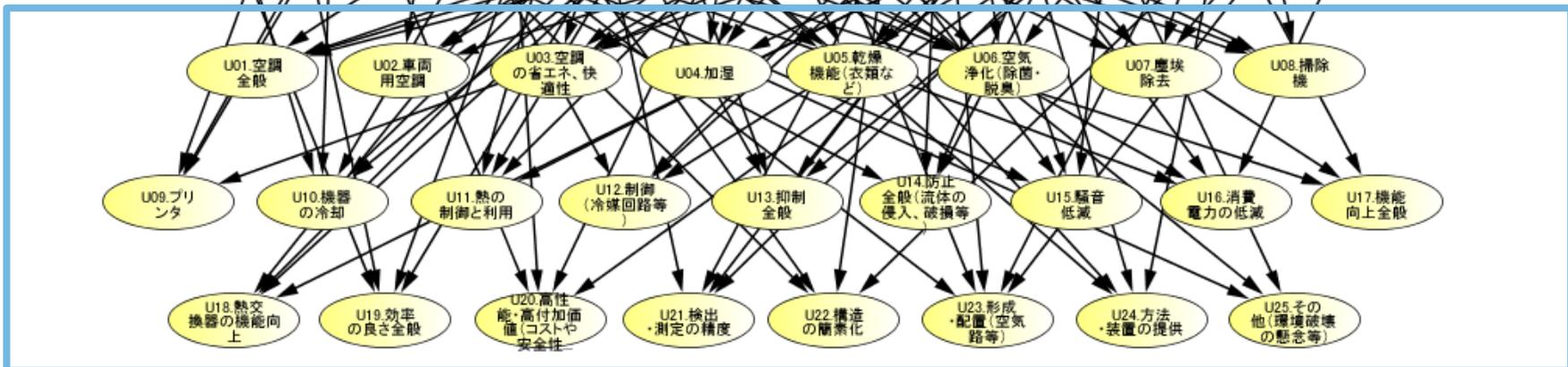
- 「T28.イオン発生・空気除菌・脱臭と「T29.電解水生成と除菌」は、それぞれG社とI社が高シェア高注力度のポジションを確立した一強状態の技術といえる
- 「T26.放電式ミスト生成」と「T30.塵埃吸込&効率性」は、高シェア高注力度のポジションは空いているが、どちらもF社がシェアは低いものの注力度が高いポジションにある
- 一強状態の技術を避けて「U06.空気浄化」の用途を実現する場合、T26やT30の技術の開発が狙い目といえるが、注力度の高いF社は要注目である
- 一強状態にあるT28やT29の技術において、その一強企業と提携あるいはM&Aを実現すれば、その技術領域ごと獲得できる

【③用途と技術の関係分析】 技術⇒用途の関係モデル

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

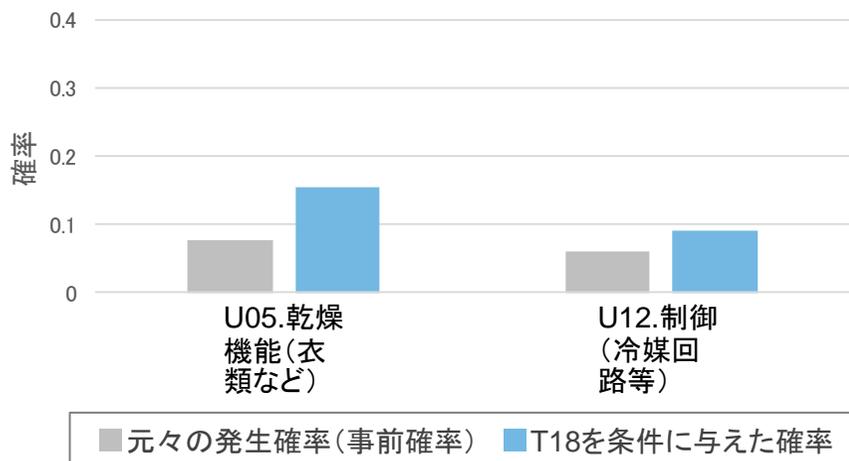


用途トピック

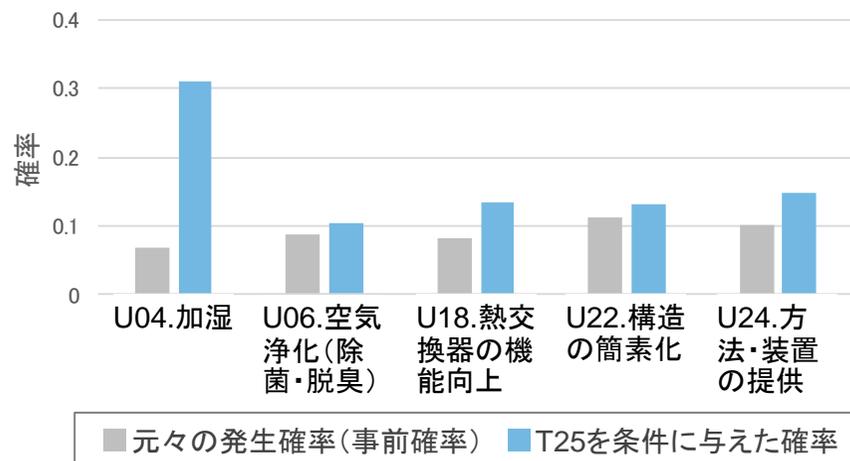
【③用途と技術の関係分析】 技術と関係のある用途の確認

1つの技術トピックを条件に与えたときに、ベイジアンネットワークで関係が見られた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

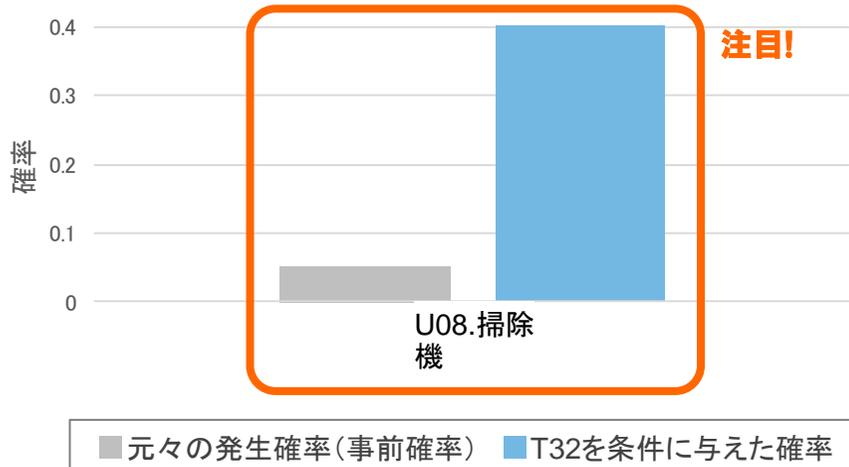
「T18.除湿」と関係のある用途



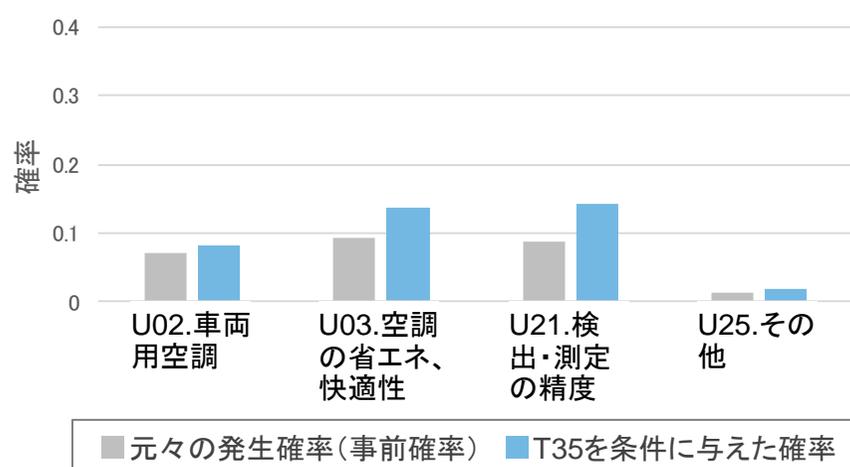
「T25.加湿」と関係のある用途



「T32.塵埃分離」と関係のある用途

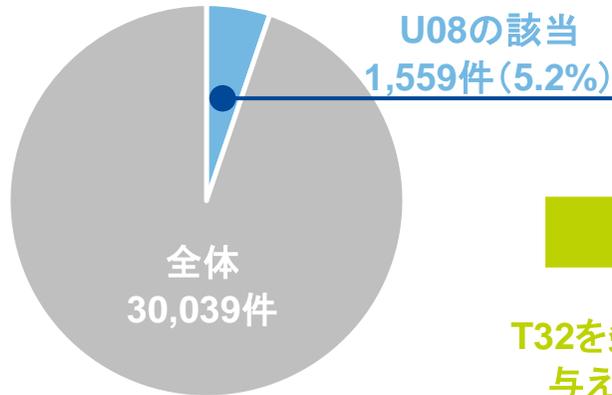


「T35.運転と停止の制御」と関係のある用途

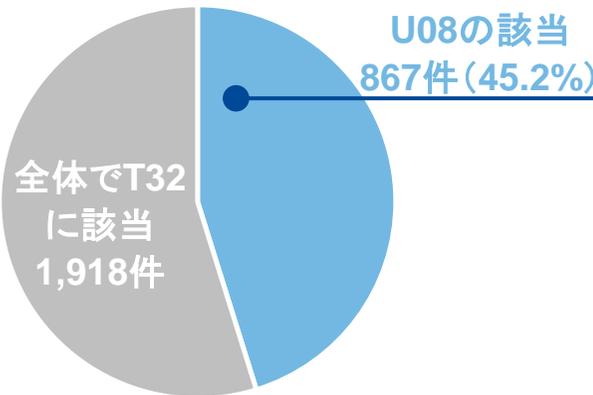


「T32.塵埃分離」の技術の応用先として「U08.掃除機」の用途はとても高い関連性がありますが、出願人Sの保有するT32ではそれが全くなく、新規用途となる可能性があります

全体でのU08の該当割合



全体でのT32におけるU08の割合



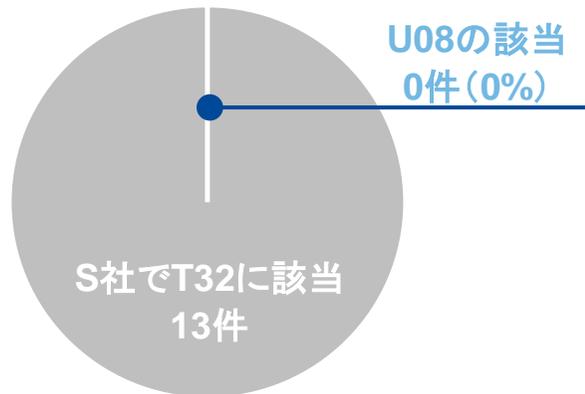
T32を条件に
与えると

T32 × U08の用途語

T32 × U08に該当する特許の【課題】文章で特徴的に出現する用途ワード

浮遊塵埃
使い勝手 プリーツフィルタ
遠心分離 吸込仕事率 分離性能
集塵容器 吸引力 捕集
旋回室
提案 廃棄
分離 掃除機 集塵部
経時的 捕集+できる
吸い込むごみ 集塵装置 本体
細い 集塵室 溜める サイクロン分離装置
別個 低下+しにくい 廃棄+できる

出願人SのT32におけるU08の割合



考察

- ベイジアンネットワークのモデルでは、技術「T32.塵埃分離」に対する用途「U08.掃除機」の関係が見られた
- 全体では、U08の該当は5.2%だが、T32を条件としたときでは、その該当割合が45.2%となりとても高い関連性が認められる
- しかし、出願人Sでは、T32に該当する特許のうち、U08に該当する特許は0件である
⇒S社の保有するT32の技術はU08の用途への展開も考えられる

【③用途と技術の関係分析】「T32.塵埃分離」の技術を「U08.掃除機」の用途で応用するアイデア

S社の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の向上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】

塵埃を含む空気を回転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】

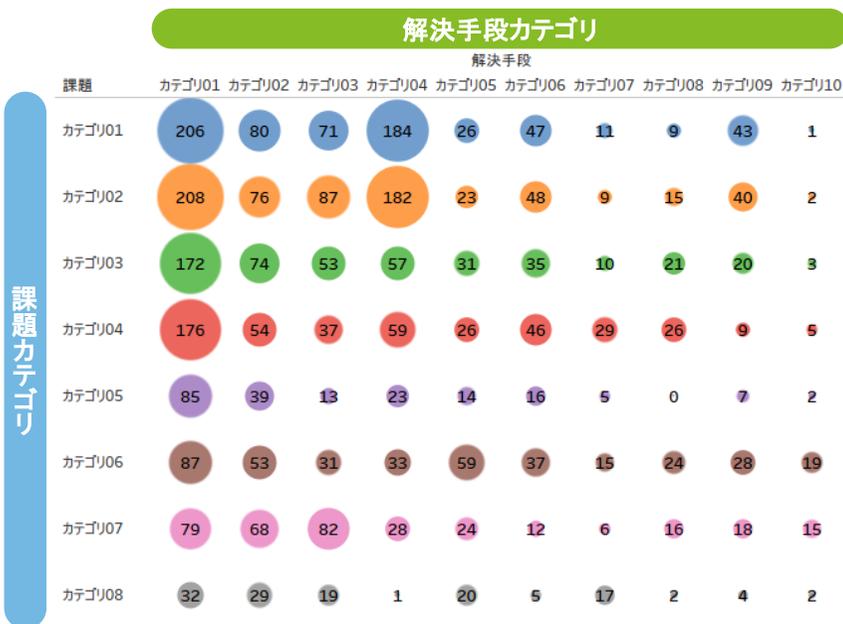
画像形成装置は、トナー含有空気からトナーを遠心分離するサイクロン部と、サイクロン部によって分離されたトナーを回収する回収部と、サイクロン部によってトナーが分離された空気を通過させ、残留トナーを捕集するフィルタ部と、空気を吸引する送風部と、フィルタの汚れを検知する汚れ検知センサが設けられたトナー捕集部を備え、汚れ検知センサで検知されたフィルタの汚れから推定した風量と、風速センサで取得した風量の実測値の差分が、サイクロン清掃閾値を超えたと判断すると、サイクロン部の清掃モードを実行する。

※対外説明用のため要約文は一部加工している

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

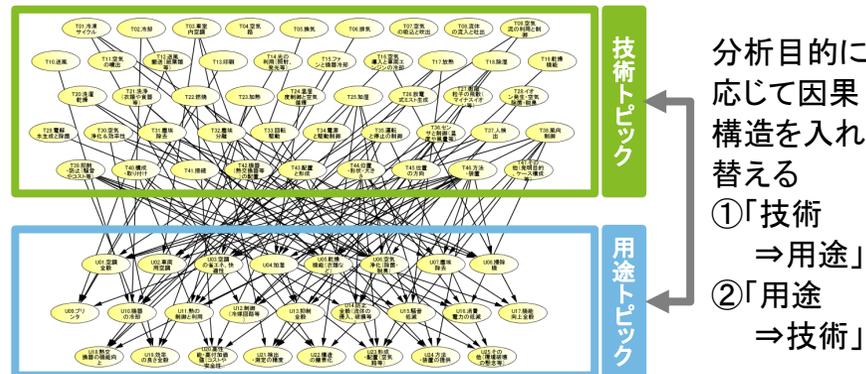
課題と解決手段のカテゴリ間のクロス集計



- 【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)

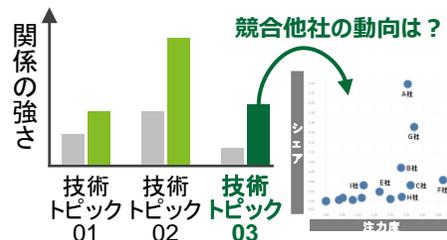
Nomolyticsを適用した特許分析

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

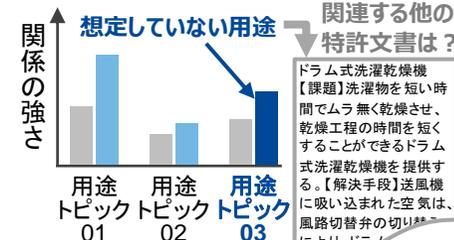


分析目的に応じて因果構造を入れ替える
 ①「技術 ⇒ 用途」
 ②「用途 ⇒ 技術」

想定用途と関係のある技術



自社技術と関係のある用途

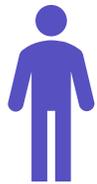


- 客観的に抽出されたトピックをベースに課題と解決手段(用途と技術)の統計的な関係をベイジアンネットワークで把握できる
- 検討中の用途に対して、関係の強い技術を確認し、各技術における出願人の動向から自社の技術戦略を検討できる
- 自社技術と関係の強い用途で想定していないものを確認し、その関連特許の探索から技術の新規用途アイデアを創出できる

PCSA と *differential* PLSA

インサイト獲得を狙ったPLSAの新展開技術

特に新たな気づきとなるインサイトの発見が求められるビジネス現場では、データ分析の結果に目新しさがなく、期待外れと評価されることがあります



この結果は経験的に
よく知ってるんだよな…



期待外れだ…



これまで見えてなかったこと
が発見できるんじゃないの？



なんだ目新しく
ないな…

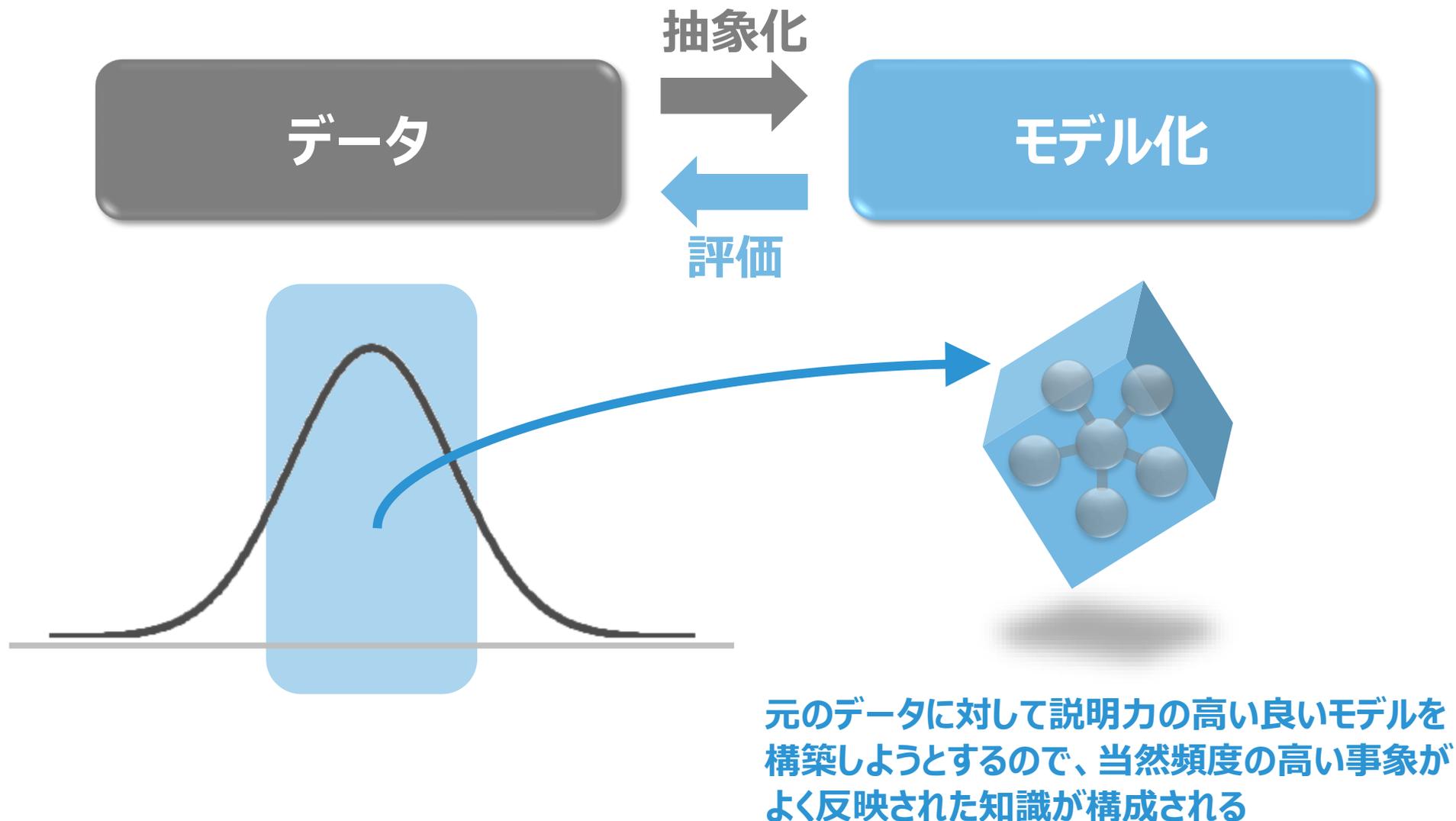


わざわざ分析しなくても
分かってるんだけど…



ビジネスはインサイト
だよ！インサイト！

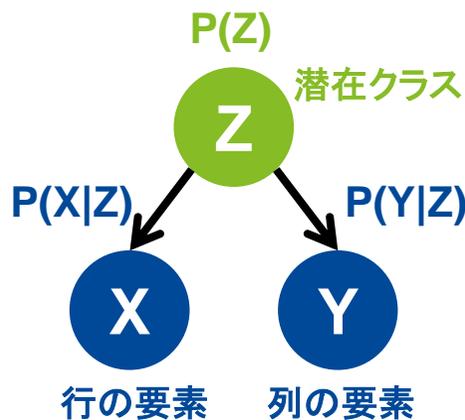
そもそもデータをモデル化することは、データに潜む傾向やルールを抽象化することで、説明力の高いモデルの構築では頻度の高い事象 \equiv 経験的によく知っていることが優先的に表現されます



テキストデータ全体を表現する代表的・典型的なトピックではなく、より特徴的・個性的なトピックを抽出するPLSAの新しい応用手法、“PCSA”と“*differential PLSA*”を開発しました



室内における人の位置を確実に把握して人に直接冷風あるいは温風が当たらないように風の到達位置を制御することで快適空調を達成することができる空気調和機を提供する。室内機に設けられたファンの回転数と、室内機に設けられた上下羽根の角度を制御して、暖房時は人体検知センサにより検知された人の足元近傍に温風を到達させる一方、冷房時は人体検知センサにより検知された人の頭上上方に冷気を到達させるように設定する。



	Y_1	Y_2	...	Y_n
X_1				
X_2				
...				
X_m				

共起行列 $n(X_i, Y_j)$

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



通常は典型的なトピックが抽出されやすいが...

より特徴的・個性的なトピックを抽出する

PCSA
differential PLSA

PCSA

ターゲットに特化したトピックを抽出してより顕著な特徴を探索する

PLSAでは、テキストマイニングで全体のテキストデータから抽出した単語で構成される共起行列をインプットにすることで、データ全体を表現する代表的なトピックを抽出します

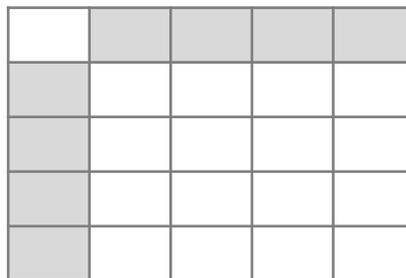


単語抽出



テキストデータにテキストマイニングを実行し、全体のデータに含まれる単語を抽出する

共起頻度集計



抽出した各単語の全体における共起頻度を集計した共起行列を作成する

PLSA

トピック抽出



作成した共起行列にPLSAを適用し、使われ方の似ている単語をトピックに集約する

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要因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)



単語抽出

共起頻度集計

トピック抽出



出現頻度

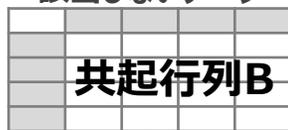
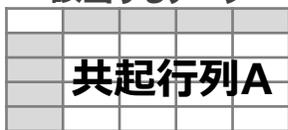
- Word 1
- Word 2
- Word 3
- Word 4
- Word 5
- Word 6
- Word 7
- Word 8
- Word 9
- Word x

あるターゲット事象(属性情報)Xが、



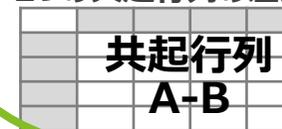
該当するデータ

該当しないデータ



2つの共起行列の差分

PLSAの適用



トピック1



トピック2



トピック3



ターゲット事象Xの該当有無に影響を与える要因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

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



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はその属性の特徴をよく表すトピックを最初から抽出し、より顕著な特徴の探索を行います

Nomolyticsでは

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では

属性の特徴をよく表現する

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

PCSAを適用した特許分析事例

電気自動車関連の特許データを用いたトレンド形成トピックの抽出

電気自動車関連の10年分の特許データ26,419件の要約全文を対象に、「出願年」をターゲットにPCSAを適用することで、近年上昇傾向あるいは下降傾向にあるトピックを抽出します

データの抽出条件と分析対象

- 対象
 - 公開特許公報
- キーワード
 - 要約と請求項に「車」と「電気」を含む
- 出願日
 - 2007年1月1日～2016年12月31日
- 抽出方法
 - Patent Integrationを使用
- 抽出件数
 - 26,419件
- 分析対象
 - 要約の全文



PCSAのターゲット



分析プロセス

テキストマイニング

Text Mining Studio

要約文から単語や係り受け表現を抽出

差分共起行列の作成

出願年が2013年以前の共起行列と2014年以後の共起行列の差分を取った共起行列を作成

トピック化 (PLSA)

Visual Mining Studio

差分共起行列にPLSAを適用して2014年前後の出願に特徴があるトピックを抽出

トピックのスコアリング

全データに対する各トピックのスコアを計算

特徴の可視化

トピックのスコアを出願年を軸に集計・可視化

【従来手法】 通常のPLSAによるトピック抽出のアプローチ

全ての要約文データから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、データ全体を表現するトピックを抽出します



要約文に含まれる「単語(名詞)」と「係り受け」を抽出する

単語	頻度
構成	4,997
制御	4,360
配置	3,895
モータ	3,486
形成	3,459
供給	3,309
検出	3,215
電気自動車	3,181
...	...

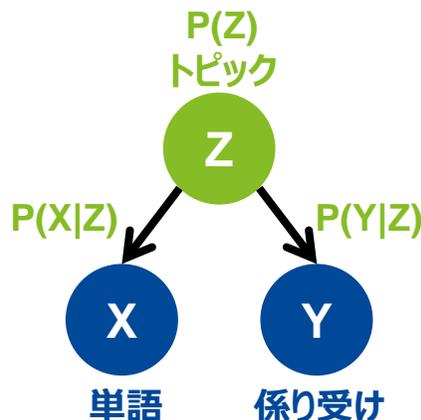
係り受け表現	頻度
電力⇒供給	1,208
否⇒判定	517
モータ⇒駆動	460
バッテリー⇒充電	440
効率⇒良い	419
供給⇒電力	332
電気自動車⇒提供	285
充電⇒行う	273
...	...

全ての要約文データを対象とした「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

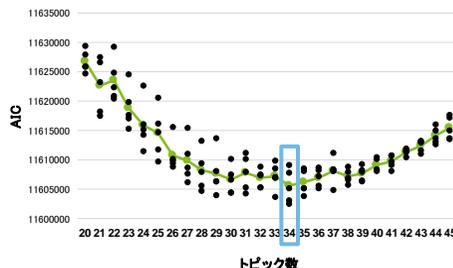
		係り受け表現				
単語		電力↓供給	否↓判定	モータ↓駆動	バッテリー↓充電	...
	構成	118	33	36	33	
制御	268	73	108	85		
配置	69	2	29	8		
モータ	239	61	494	58		
...						

単語(名詞): 3,020語
係り受け: 2,128表現
※頻度20件以上を対象

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 $P(Z)$
トピックの存在確率
 - ② $P(X|Z)$
トピックにおける単語の所属確率
 - ③ $P(Y|Z)$
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 $P(X|Z)$ と $P(Y|Z)$ からトピックの意味を解釈する

トピック Z13	
$P(Z) = 5.0\%$	

$P(X Z)$	単語	$P(Y Z)$	係り受け
12.6%	充電	5.1%	バッテリー-充電
8.9%	電気自動車	4.0%	充電-行う
6.5%	蓄電装置	3.9%	電気自動車-充電
3.0%	バッテリー	1.9%	蓄電池-充電
2.0%	充電システム	1.6%	蓄電装置-充電
2.0%	蓄電池	1.6%	電力-供給
1.9%	電力	1.3%	充電-開始
1.7%	制御	1.2%	電気自動車-接続
1.5%	充電スタンド	1.2%	充電-蓄電装置
1.5%	放電	1.1%	用いる-充電
...

確率の高い構成要素から、トピック Z13は「電気自動車の蓄電池充電」に関するトピックと解釈できる

通常のPLSAによる全体トピック34個の一覧②

26,419件の特許は、エンジン、動力伝達、モータ、ブレーキ、電力変換、二次電池、充電、情報通信、異常検出、筐体、構成、小型化、安全性などの34個のトピックに集約されました



Z32,Z33,Z34は特定の出願人による重複した要約内容の特許から抽出された

※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（関係の強い上位5つの単語を赤色で表示している）

【開発手法】PCSAによるトピック抽出のアプローチ

① 2013年以前データ、② 2014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、2014年前後の出願に特徴があるトピックを優先的に抽出します

2つの共起行列の作成

①出願年が**2013年以前**のデータ
20,014件(文章数:33,113件)

係り受け表現

	電力↓供給	否↓判定	モータ↓駆動	バッテリー↓充電	...
構成	97	30	27	25	
制御	208	49	80	66	
配置	56	1	23	8	
モータ	192	42	356	46	
...					

単語

②出願年が**2014年以後**のデータ
6,405件(文章数:10,723件)

係り受け表現

	電力↓供給	否↓判定	モータ↓駆動	バッテリー↓充電	...
構成	21	3	9	8	
制御	60	24	28	19	
配置	13	1	6	0	
モータ	47	19	138	12	
...					

単語

差分共起行列の作成

①2013年以前の共起行列と
②2014年以後の共起行列の
差の絶対値を計算した共起行列
を作成する

係り受け表現

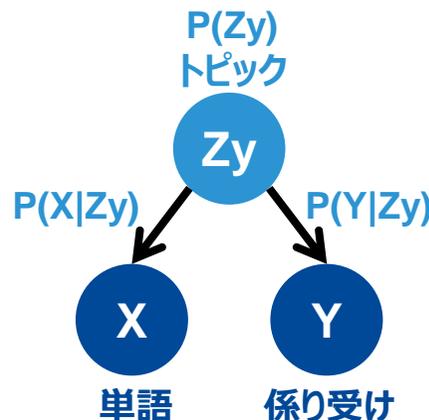
	電力↓供給	否↓判定	モータ↓駆動	バッテリー↓充電	...
構成	10.4	6.7	0.3	0.1	
制御	7.4	8.1	2.1	2.4	
配置	5.1	0.7	1.4	2.6	
モータ	15.2	5.4	22.7	2.9	
...					

単語

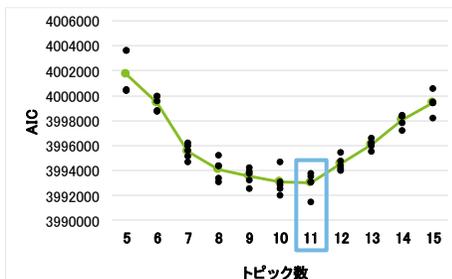
ただし、2つの共起行列は異なる
文章数のデータから作成されて
いるので、①2013年以前の共起
行列の頻度を2つの文章数の比
率(10,723/33,113)で重み調整
してから、②2014年以後の共起
行列との差を計算する

PLSAの実行

差分共起行列に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 $P(Z_y)$
トピックの存在確率
- ② $P(X|Z_y)$
トピックにおける単語の所属確率
- ③ $P(Y|Z_y)$
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 $P(X|Z_y)$ と $P(Y|Z_y)$
からトピックの意味を解釈する

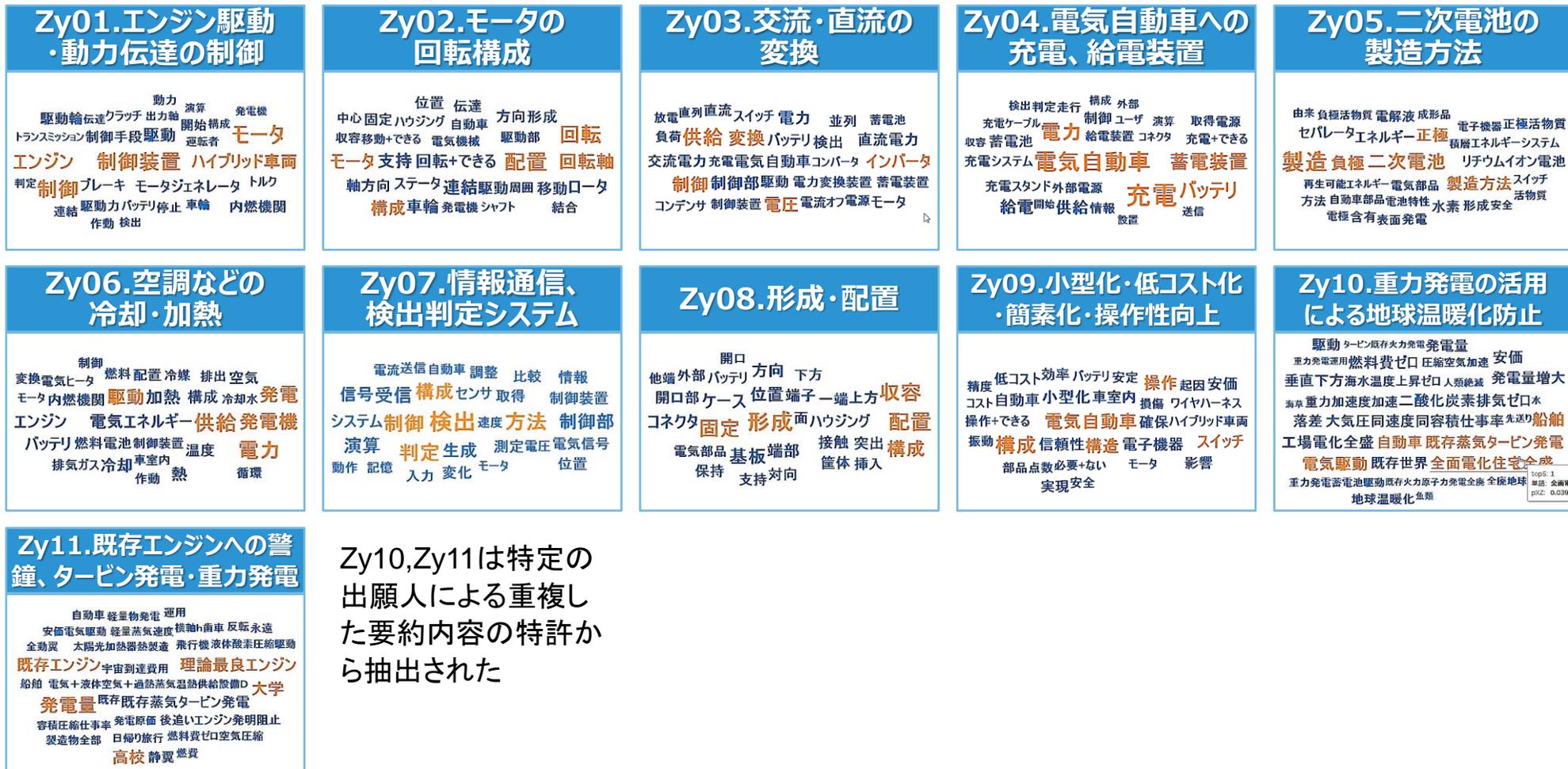
トピック Z_y05

$P(Z_y) = 2.5\%$

$P(X Z_y)$	単語	$P(Y Z_y)$	係り受け
3.1%	製造	4.5%	水素-製造
2.2%	二次電池	3.0%	発電-電気
1.5%	負極	2.6%	製造方法-提供
1.5%	正極	2.6%	製造-方法
1.5%	製造方法	1.6%	含む-リチウムイオン電池
1.2%	エネルギー	1.3%	成形品-提供
1.1%	リチウムイオン電池	1.3%	強度-有する
1.1%	水素	1.3%	表面-形成
1.1%	セパレータ	1.2%	含む-組成物
1.0%	電解液	1.2%	リレースイッチ
...

確率の高い構成要素から、トピック
Z_y05は「二次電池の製造方法」に
関するトピックと解釈できる

出願年(2014年前後)で特徴を示すトピックは、エンジン駆動、モータ構成、電力変換、充電、二次電池の製造方法、冷却・加熱、検出判定、小型化・低コスト化など11個抽出されました



Zy10, Zy11は特定の出願人による重複した要約内容の特許から抽出された

※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 (関係の強い上位5つの単語を赤色で表示している)

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)を計算し、各トピックが該当するデータのうち、“2014年以後”となる割合について、PLSAの結果とPCSAの結果を比較します

PLSAによるトピックとPCSAによるトピックのスコア(フラグ情報)を紐づけた特許データ

PLSAによる34個の全体トピック

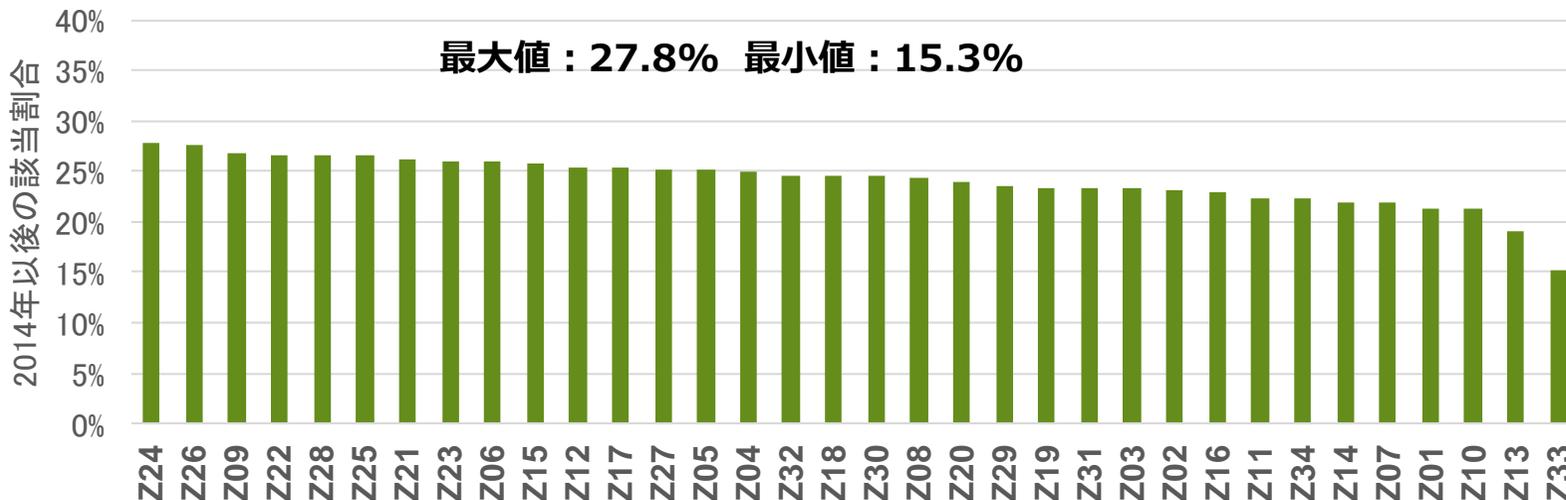
PCSAによる11個の要因トピック

特許ID	出願番号	出願年	出願年グループ	トピック Z01	トピック Z02	...	トピック Z34	トピック Zy01	トピック Zy02	...	トピック Zy11
1	特願2007-XXXX	2007	①2013年以前	1	1		0	0	1		1
2	特願2009-XXXX	2009	①2013年以前	0	1		1	0	0		1
3	特願2012-XXXX	2012	①2013年以前	0	1		1	0	1		0
4	特願2014-XXXX	2014	②2014年以後	1	0		0	1	0		0
...
26,419	特願2016-XXXX	2016	②2014年以後	1	0		1	1	0		0

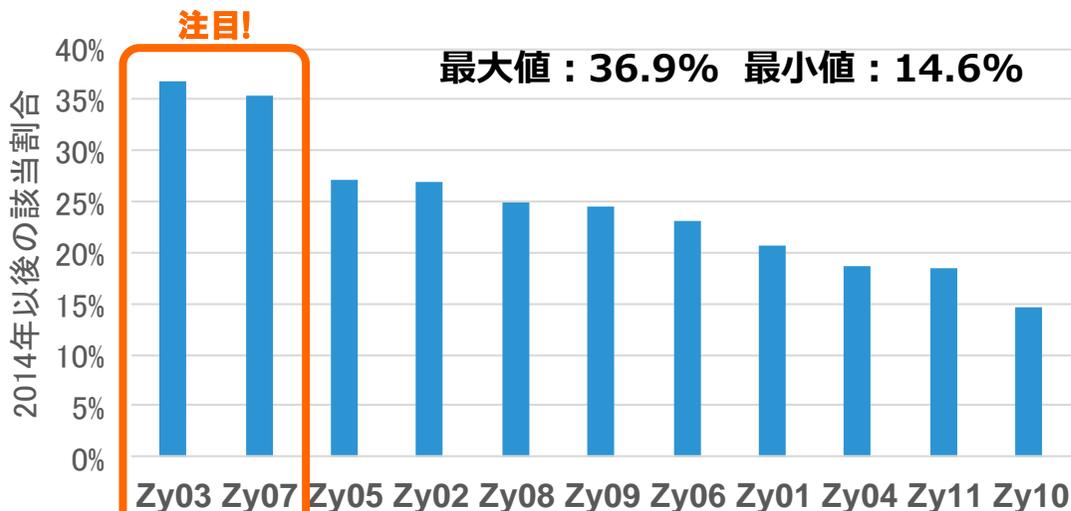
各トピックが該当する特許のうち、出願年が「②2014年以後」の割合を比較することで、2014年前後においてPCSAではどれくらい偏ったトピックを抽出できているか確認する

各トピックの2014年以後の割合は、PLSAではおおむね25%前後ですが、PCSAでは割合が高いものと低いものに偏っており、2014年前後に対して特徴的なトピックとなっています

PLSA
による
全体
トピック



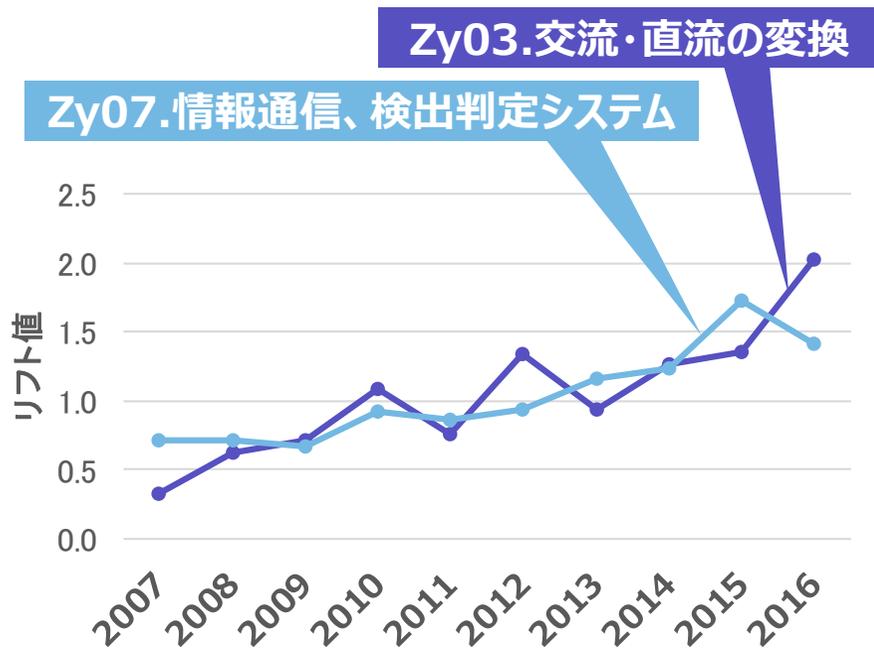
PCSA
による
要因
トピック



- それぞれ2014年以後の該当割合の高い順にトピックを並べている
- PLSAではおおむね25%前後のトピックが抽出されているが、PCSAのトピックでは割合が高いものから低いものまで抽出されており、最大値も高い
- (参考)元々の2014年以後データの割合は24.2%

2014年以後の割合が高いトピックとして、電力変換に関する制御技術(Zy03)、データに基づいた運転者のアシスト技術(Zy07)が近年上昇傾向にあることが分かります

2014年以後の割合がトップ2のトレンド



$$\text{※リフト値} = \frac{P(\text{出願年} | \text{トピックTy}=1)}{P(\text{出願年})}$$

- 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に対して、その出願年の出願割合で正規化した値であり、全体における各出願年の出願割合がそのトピックを条件にすることで何倍に変化するかを示す

該当特許の要約の例

Zy03.交流・直流の変換

発明の名称	出願年
電気自動車	2016

要約文 (抜粋)

車両の衝突時に平滑化コンデンサを放電する確実性を向上させる。衝突検知装置が、衝突を検知したときに第1信号と第1信号に続く第2信号を送信する。第1信号を受信したときに、インバータ制御回路がスイッチング素子駆動回路への電力の供給を停止する。第2信号を受信したときに、インバータ制御回路が平滑化コンデンサを放電する。

Zy07.情報通信、検出判定システム

発明の名称	出願年
車両速度の制御方法	2016

要約文 (抜粋)

経路及び交通状況に関する情報の応答として、自動車の速度を制御する方法に関するものである。本方法は、計画経路データ及び/または、繰返行程ロガーデータにより特定された想定経路に基づいて、最適な制動または加速点を決定し、最適な制動または加速点に基づいて、運転者に速度プロファイルを調整するためのサインを送る。

Nomolyticsはデータ全体を表すトピックを理解し、その特徴を様々な分析軸で探索し、PCSAは探索したい特徴に特化したトピックを抽出し、より効果的なインサイト獲得を狙います

Nomolytics®

テキストデータ全体を表すトピックを理解し、その代表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

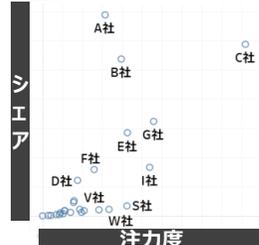
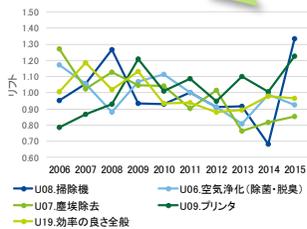
ベイジアン
ネットワーク

単語抽出

トピック化

モデリング

全体を表現するトピックとは？



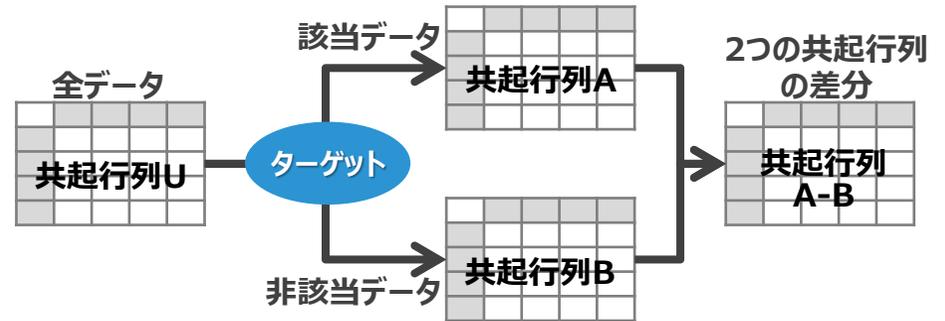
PCSA®

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

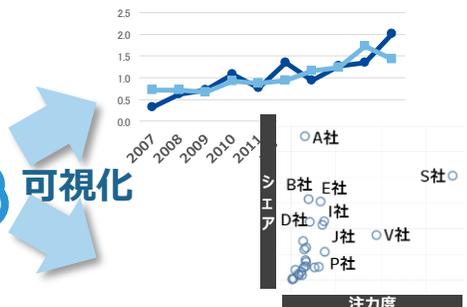
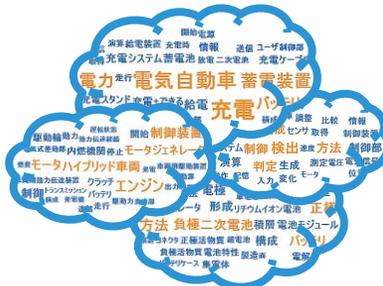
テキスト
マイニング

差分の
共起行列

PLSA



ターゲットの該当有無を左右するトピックとは？

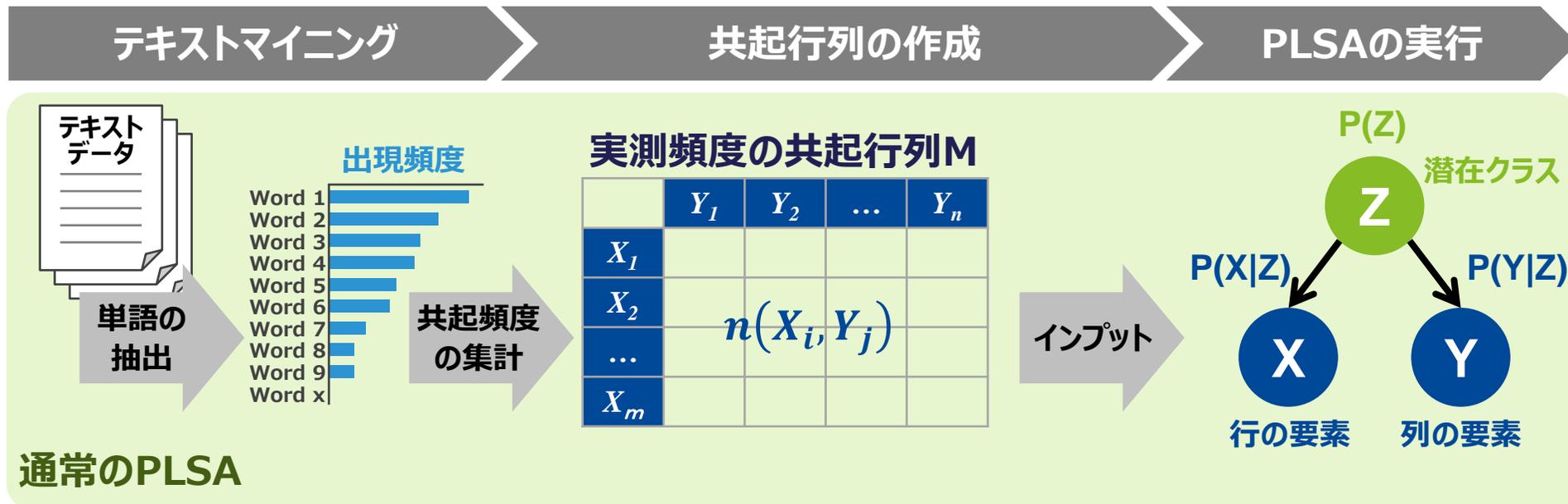


differential PLSA

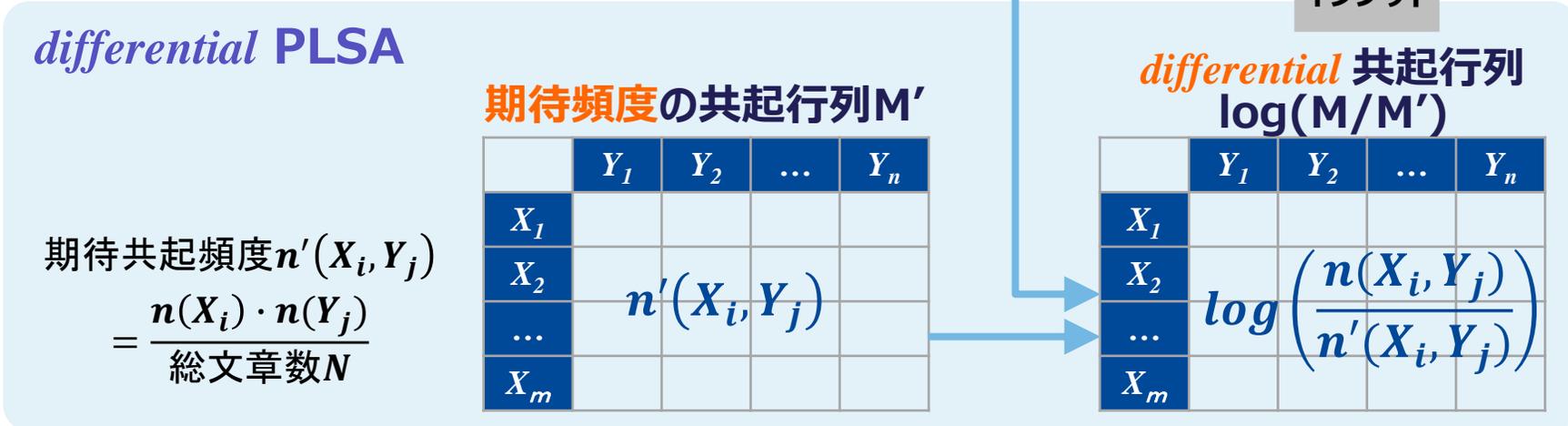
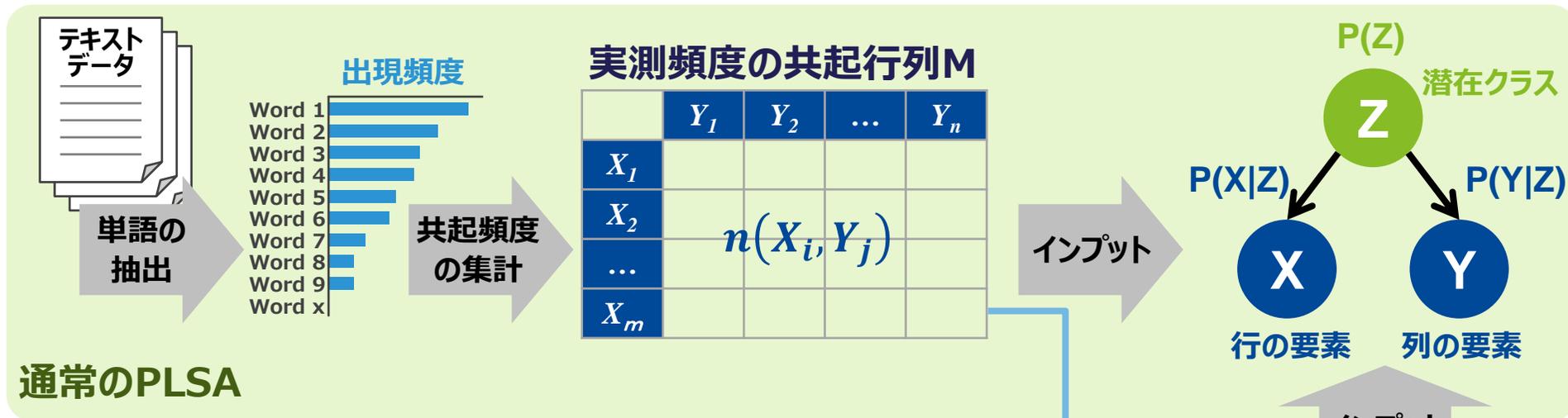
頻度によらない個性的なトピックを抽出して埋もれがちな特徴を発見する

【従来手法】PLSAを用いたトピックの抽出

通常のPLSAでは、テキストデータに含まれる単語を抽出し、その単語群の共起頻度を集計した共起行列をPLSAのインプットとしてトピックを抽出します



differential PLSAでは実測頻度を集計した共起行列に加え、期待頻度を集計した共起行列を作成し、その比率の対数を取った共起行列をPLSAのインプットとしてトピックを抽出します



通常のPLSAでは、頻度の高い要素が優先的に反映され典型的なトピックになりがちですが、*diff*-PLSAでは、頻度が低い要素も存在度が増し、個性的なトピックの抽出が期待できます

通常のPLSA

- PLSAを適用した確率解では、**頻度が高い要素に高い確率が割り当てられる**傾向にある
- 結果として抽出されるトピックは**典型的なもの**になってしまう

differential PLSA

- 実測共起頻度を期待共起頻度で除した値の対数を値とする共起行列をインプットとする
- 実測共起頻度が高い共起ペアでも、元々全体の頻度が高ければ期待共起頻度も高くなり、その比率を取ることで値の大きさが制限される
- 逆に実測共起頻度が低い共起ペアでも、期待共起頻度がそれよりも十分低ければ比率の値は大きくなり、PLSAの解ではこうした**頻度の低い要素にも高い確率が割り当てられる**可能性がある
- 結果として、**より個性的なトピック**が抽出されることが期待できる
- 単純な比率ではなくその対数を取ることは、期待頻度が低すぎることで極端に高くなる値を制限し、必要以上にデフォルトされた歪んだトピックが抽出されることを防ぐ効果がある

differential PLSAを適用した特許分析事例

電気自動車関連の特許データを用いた個性的なトピックの抽出

先ほどと同様の電気自動車関連の特許データ26,419件の要約全文を対象に、*differential PLSA*を適用することでより個性的なトピックを抽出します

データの抽出条件と分析対象(先ほどと同様)

- 対象
 - 国内の公開特許公報
- キーワード
 - 要約と請求項に「車」と「電気」を含む
- 出願日
 - 2007年1月1日～2016年12月31日
- 抽出方法
 - Patent Integrationを使用
- 抽出件数
 - 26,419件
- 分析対象
 - 要約の全文



分析プロセス

テキストマイニング

Text Mining Studio

要約文から単語や係り受け表現を抽出

差分共起行列の作成

実測頻度の共起行列から期待頻度の共起行列を計算し、両者の比率の対数を取った共起行列を作成

トピック化 (PLSA)

Visual Mining Studio

差分共起行列にPLSAを適用してより個性的なトピックを抽出

トピックのスコアリング

全データに対する各トピックのスコアを計算

特徴の可視化

トピックのスコアを出願年や出願人を軸に集計・可視化

①実測頻度の共起行列と②期待頻度の共起行列の比率の対数を値とする差分共起行列にPLSAを適用することで、より個性的な特許トピックを抽出します

実測と期待の共起行列作成

①実測頻度の共起行列

		係り受け表現				
		総頻度 n(Y)	1,350	539	494	529
単語	総頻度 n(X)		電力↓供給	否↓判定	モータ↓駆動	バッテリー↓充電
	5,880	構成	118	33	36	33
	5,092	制御	268	73	108	85
	4,604	配置	69	2	29	8
	5,188	モータ	239	61	494	58

②期待頻度の共起行列

		係り受け表現				
		総頻度 n(Y)	1,350	539	494	529
単語	総頻度 n(X)		電力↓供給	否↓判定	モータ↓駆動	バッテリー↓充電
	5,880	構成	34.6	13.8	12.7	13.5
	5,092	制御	29.9	12.0	11.0	11.7
	4,604	配置	27.1	10.8	9.9	10.6
	5,188	モータ	30.5	12.2	11.2	12.0

※総文章数: 229,598件

差分共起行列の作成

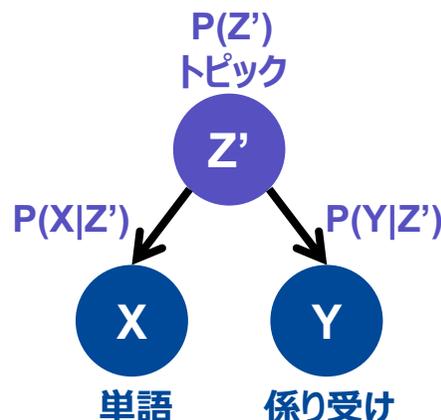
2つの共起行列の各共起ペアにおいて、期待共起頻度に対する実測共起頻度の比率の対数を取った差分共起行列を作成する

		係り受け表現			
		電力↓供給	否↓判定	モータ↓駆動	バッテリー↓充電
構成	1.2	0.9	1.0	0.9	
制御	2.2	1.8	2.3	2.0	
配置	0.9	-1.7	1.1	-0.3	
モータ	2.1	1.6	3.8	1.6	

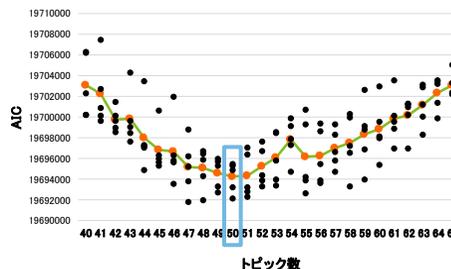
※値が負数となるものは"0"に置換する

PLSAの実行

差分共起行列に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 $P(Z')$
トピックの存在確率
 - ② $P(X|Z')$
トピックにおける単語の所属確率
 - ③ $P(Y|Z')$
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 $P(X|Z')$ と $P(Y|Z')$ からトピックの意味を解釈する

トピック Z'20			
P(Z') = 2.3%			
P(X Z')	単語	P(Y Z')	係り受け
2.2%	プラグ	1.3%	充電ケーブル-接続
1.8%	コンセント	1.2%	プラグ-接続
1.7%	インレット	1.2%	外部電源-接続
1.6%	充電ケーブル	1.1%	コネクタ-介する
1.6%	充電プラグ	1.1%	充電コネクタ-接続
1.5%	外部電源	1.1%	充電プラグ-接続
1.3%	充電コネクタ	1.0%	コネクタ-接続
1.3%	電源プラグ	0.9%	接続-状態
1.2%	接続+できる	0.9%	供給-蓄電装置
1.0%	接続部	0.9%	接続+できる-構成

確率の高い構成要素から、トピック Z'20は「充電の接続」に関するトピックと解釈できる

diff-PLSAでは、通常のPLSAのトピックに加え、運転者の補助、発電・蓄電、燃料電池、情報の取得・提供、車両用灯具、組成物の製造などを含む50個のトピックが抽出されました

Z'01.エンジン制御

オン
始動要求 要求 判定 モード差 時点
オフ 所定条件 成立条件 電気走行モード プログラム
所定時間 解放 ハイブリッドECU所定値モータジェネレータ
経過 再始動 充電状態継続 ハイブリッド車両
エンジン ECU実行中 禁止 NO 停止 開始

Z'02.動力伝達

変速機構
出力要素 出力軸 連結できる 出力部材 電気機械 適用
動力逆並車装置 駆動軸 遊星歯車機構 キャリア
入力部材 トルク 動力伝達+できる ギヤ トランスミッション
連結 入力軸 回転駆動+できる 伝達 サンギヤ 伝達+できる
リングギヤ クラッチ 内燃機関 変速比 動力伝達機構
動力伝達駆動輪

Z'03.差動機構などを備えた動力伝達の制御

変速制御運転状態 差動部 変速比
駆動源無段変速機 動力伝達装置 高速回転 自動変速部
駆動輪車両用駆動装置 回転部材 ハイブリッド走行モード 自動変速機
動力伝達経路 車両用動力伝達装置 エンジン回転速度
エンジン モータジェネレータ 変速ショック 電気式差動部
電気走行モード 差動状態 制御装置 変速部 変速
エンジン始動時 エンジン始動 変速時 ショック

Z'04.回転運動

駆動軸 同軸
伝達機構 軸方向外周面 回転カ一方
ハブ 回転軸 シャフト 羽根車 ロータ 中心 直線運動
クランクシャフト 磁石 回転運動 チェーン 相対回転+できる
角度位置 回転体 回転 回転+できる 回転駆動
回転方向 伝達 ステータ ギヤ軸 反対側

Z'05.ロータ・ステータなどモータの構成

シャフト 巻装
スロット 整流子 回転磁界 外周 径方向
極 界磁巻線回転+できる コイル ヨーク 電機子 プラシ
永久磁石 巻線 車両用交流発電機 軸方向 固定子巻線
固定子鉄心 回転軸整流器 ステータ プラシホルダ
内周 ロータ周方向磁極 摺 整流

Z'06.モータ制御(トルク制御や回転数制御など)

制御トルク制御 回転数トルク推定 モータ一致
車速 回転角度 出力トルク モータ制御装置 電気角
トルク指令値 位相トルク指令 回転位置 モータ制御部
モータトルク 電圧指令 電流指令演算 スイッチング素子
インバータ 回転速度 誘起電圧加算電流 補正 絶対値
振幅

Z'07.油圧ポンプなどを利用したモータ駆動

電力 供給 車輪
駆動 エンジン 動力トルク 走行用後輪
走行トランスミッション ハイブリッド車両 再生電力
走行用モータ 発電 油圧ポンプ 発電電力 電気エネルギー
発電機 油圧モータ 駆動輪 回転数 モータジェネレータ
動力源 内燃機関 駆動力伝達 圧油 モータ
前輪

Z'08.ブレーキ

押圧力操作部材 ピストン制御操作部材 摩擦部材 リザーバ 運転者付与
車輪 モータ駆動装置 制動力 プレーキ操作 液圧
スレーブシリンダ プレーキ液圧 入力装置 マスタシリンダ
ブレーキ 電気ブレーキ プレーキ液圧 駆動トルク プレーキペダル
ブレーキアクチュエータ アクセルペダル 操作力 ホールシリンダ
操作量 回転部材操作者 操作

Z'09.状態に応じた制御、運転者の操作補助

点検 移行 停止 操作 作動 禁止
運転者閉状態 検出+ない 検出部 検出結果 アンロック
ソフトレバー処理 シフトレンジ パーキングレンジ
ロック 絶縁抵抗 制御手段運転席オン 解除 キースイッチ
開状態 自動的 検出 警告 切替故障
開始

Z'10.コンバータとバッテリー昇降圧

給電 出力電圧
コンデンサ 主電源 再生電力並列 昇圧 放電
供給 電気負荷 低電圧バッテリー補機 バッテリーサブバッテリー
電圧コンバータ DCコンバータ 降圧 車両用電源装置
高電圧バッテリー鉛蓄電池 電気二重層コンデンサ
電源システムオルタネータ 電源装置 電圧コンバータ
バッテリー 発電機 電力 直流

Z'11.直流と交流の電力変換

直流電力変換部 コンバータ 交流電圧
主電動機 フィルタコンデンサ 直流電源 架線 スイッチング動作
交流電力 変圧 電気車制御装置補助電源 直流電力
直流電圧 集電装置昇圧 電力変換装置 交流電圧
交流電動機 インバータ 再生電力 電圧 交流 供給
変換電力変換

Z'12.回転力などの電気エネルギー変換

海水
水車 太陽光 上下動 流体 蓄電
風運動エネルギー 風車 回転 ナセル発電システム
機械エネルギー 水 回転エネルギー 回転軸 電気エネルギー
波 回転力 発電機 タービン 風力発電装置 エネルギー
水流 変換排気ガス発電 水面 羽根

Z'13.エネルギー効率の向上

快進損失 技術 走行+できる 精度 回避
燃費 電気自動車 無駄乗員 確保 走行距離
長寿命化 効率エネルギー 効率劣化 ハイブリッド車両
高速走行時 航続距離 短縮 燃料消費量 消費電力
電力消費 走行中 違和感 起因 性能 最小限

Z'14.発電と蓄電

回収
自由 蓄電 電力量 蓄電量 発光 蓄電池
電力風力発電蓄電+できる 発電夜間 運動エネルギー
電気エネルギー 再生可能エネルギー 発電電流 ソーラーパネル
太陽電池 化石燃料消費 風力発電装置 充電機能
エネルギー源 発電+できる 活用 走行 回転力稼働
照明 車両外部

Z'15.電池モジュールの提供

内蔵 構造 電極 相互
積層方向 単電池 バッテリー ベースプレート 横方向
隣接スタック 電池システム 電池モジュール 列 充放電+できる
冷却 ユニット 冷却剤 中型 積層 垂直 電池パック 底部
配列 パスワード 片側 構築 直列外形

Z'16.燃料電池

酸素 漏洩 空気排出 反応 電力
排気ガス 燃料タンク 燃料電池 加水 水素
酸化ガス 燃料電池システム 電気化学反応
燃料ガス 電気分解 内燃機関燃焼室 燃料電池車両
供給+できる 排気通路 供給ガス 浄化 貯蔵 発電
燃焼 温度

Z'17.二次電池の構成

外装部材
含窒リチウムイオン バッテリーケース 存在方向 層構造 巻回
電解質 集電体 二次電池 電極 正極活物質 電解液
非水電解質電池 表面 セパレーター 活物質リチウムイオン電池
両面 正極 負極 電解要素 積層 負極活物質
電気抵抗 負極集電体 外装体

Z'18.バッテリーの充放電

充電 電源 警報建物 放電 給電
給電装置 蓄電池 授受 充放電装置 系統電源
負荷電力変換装置 充電制御部 時間帯 充電システム
停車時 指示 電力供給システム プレーカ 放電装置
電源電圧 充放電分電盤 双方向車載 電力システム 停電
コネクタ住宅 送受信 電力 開閉器

Z'19.充電システム

充電
充電電流 駐車場 充電動作 駐車 利用者外部電源
記号 充電電圧 充電+できる 蓄電池 充電制御装置 蓄電装置
充電スタンド 充電時 充電制御部 時間帯 充電システム
電気自動車 商用電源電力量 充電量 充電制御
終了 集合住宅 充電電力電気料金情報 ユーザ
管理

Z'20.充電の接続

充電電線
電源ケーブル 商用電源 充電時 電力変換部
車両コネクタ インレット 接続部 充電口 スタンド電源プラグ
外部電源 コネクタ 充電ケーブル コンセント 充電プラグ
着脱 接続+できる 電力ケーブル 電力 蓄電装置 充電コネクタ
外部機器 プラグ 車両外部ケーブル 接続装置 外部
給電部

diff-PLSAでは、通常のPLSAのトピックに加え、運転者の補助、発電・蓄電、燃料電池、情報の取得・提供、車両用灯具、組成物の製造などを含む50個のトピックが抽出されました

<h3>Z'21.非接触など受給電装置</h3> <p>電力 電化電磁誘導 給電回路 給電部 給電装置 非接触給電装置 給電部 給電装置 送電送電コイル 非接触受電コイル 給電+できる受電部 受電部 給電受電装置 給電スペース</p>	<h3>Z'22.車両用空調など熱交換</h3> <p>蒸発 加熱 空気 排出 発熱 圧縮機 ヒータコア 冷媒 熱 冷却 温度冷却装置 ラジエータ圧縮 車両用空調装置 熱交換 電気ヒータ 流量 車室内蒸発器 熱交換器 熱源 流通 冷却水 熱媒体 暖房 循環 ヒータ 放熱 送風</p>	<h3>Z'23.冷却装置と放熱</h3> <p>流入 収容 逆止弁 循環 閉塞 吸気口 筐体 吸込口 撮像装置 収容部 走行風 排出 冷却装置 冷却システム 送風機 流入口 放出 冷却風 排出口 吐出口 所定方向 ケース内 排気口 放射状 仕切壁 排気管 外気 開口部 冷却 送風</p>	<h3>Z'24.信号の入出力と検出</h3> <p>指示信号 センサ 入力 オン 検出 情報 操作 マイコン 出力信号 CPU ステアリングシャフト 電気信号 出力+ない 処理部 検出信号 制御部 検出手段 発信 検出結果 制御信号 光信号 送信 スイッチ手段 回転角度 異常 制御回路有無 警告音 受信</p>	<h3>Z'25.電気信号の取得と変換(センサ検出など)</h3> <p>乗員超音波 撮影演算判定 距離 点灯 検出値 判定結果 物体 電気信号 受光素子 加速度センサ 表示 圧力センサ 合成 撮像素子 検出手段 受信部 受光圧力 車両制御ユニット 静電容量 衝突 点滅 消灯 検出 記憶 周期 有無</p>
<h3>Z'26.電流・電圧の検出</h3> <p>オン 印加 比較 直列 基準電圧 電流値 蓄電素子 地絡スイッチ 電圧検出器 異常平滑コンデンサ 制御回路 電流検出器 電流センサ リレー インバータ回路 電流経路 スイッチング素子 オフ 充放電電流 直流電源 電圧 電圧センサ 半導体スイッチ 短絡故障 電流検出 コンバータ 交流電圧</p>	<h3>Z'27.温度、電流、充電量などの検出と制御</h3> <p>推定 取得 検出結果 発電電圧 判定手段 温度 充電量 積算充電状態 検出手段 充電率 演算 電気負荷 判定結果 残量 電圧検出手段 充放電 検出算出手段 放電 蓄電量 電流量 電流値 バッテリ制御装置 記憶手段 制御手段 発電電力 関係 充 二次電池</p>	<h3>Z'28.演算や推定、測定などのステップを含む方法</h3> <p>加算段階 時間 演算比較 方法 推定 記憶 閾値 起動 関数 定義 推定値 測定値 パラメータ 差 再充電 充電状態 値 終了測定 記録取得 判定 最大 運転状況 更新 レベル 基準 セル</p>	<h3>Z'29.情報の取得・提供(位置情報やバッテリー残量等)</h3> <p>特定 出発地車載機器 送信 受信 検索ユーザ 残量 地図情報 目的地 記憶部 サーバ 記憶 無線通信部 充電スタンド ナビゲーション装置 表示部 位置情報 無線通信 情報 検出情報 バッテリ残量 車両情報 携帯端末 利用者 現在地表示 通信 取得 通知</p>	<h3>Z'30.スイッチなど操作装置</h3> <p>接触 破断 入力装置 装着 保持+できる 当接 ストンプラップ 形成+できる 操作部 操作体 スイッチ 固定接点 破損 操作+できる 車室内 スイッチ接点 傾き 電子機器 配線基板 操作段差 作成 可動接点 浸入 操作時 変形 安価 勝手 カ 誤操作</p>
<h3>Z'31.車両用灯具</h3> <p>反射 貫通 近接 LED ねじ クランプ 接続ハウジング 射出 発熱体 車両用灯具 灯室内 ステアリングコラム 半導体発光素子 ステアリングシャフト 基板 光源 内装 給電部材 ランプ 照射 リフレクタ 透過 光 放熱部材 ハルブ 入射 固定部 表面 基板</p>	<h3>Z'32.掃除機</h3> <p>車輪連通 前方 開口 下面 前面 回転 上方 床面 本体 ケース 排気口 吸気口 下方 回転 両端 左右一対 吸込口 塵埃 電気掃除機 電動送風機 下ケース 着脱+できる ホース 被清掃面 集塵部 吸引 ハンドル 上面 本体 後方</p>	<h3>Z'33.基板の構成</h3> <p>発光部 光源 露出 発光 面 導電 裏側 接合部 配線表示装置 検出部 基板 発光ダイオード 放熱フィン 半導体素子 車両用照明装置 表面 発光素子 光 センサ素子 はんだ付け ケース内 パターン 被検出体 ホールIC 照明装置 厚さ方向 反対側 電極 筐体</p>	<h3>Z34.回路の接続(電力変換回路など)</h3> <p>並列 トランジスタダイオード 電圧源一端 端子変換器 他端 出力端子 コンデンサコイル 各相 インダクタンス スイッチング素子 バス 一次コイル 相スイッチ 二次コイル 電流 カソード トランス 整流器 入力端子 アノード 電源スイッチ インダクタ 抵抗拘束 直列</p>	<h3>Z'35.端子接続</h3> <p>配線 電装品 嵌合 ホルダ 貫通 絶縁性 接続部材 放射状 本 電線 パスパー 雄端子 端子 嵌合+できる ワイヤハーネス 雌端子 露出 導電部材 金属製 窓ガラス 他端 導体 接地コネクタ 突出 圧接一端 挿通 両端 端部</p>
<h3>Z'36.部品・装置の収容ケース・筐体</h3> <p>ケース 挿通 フランジ 基板 露出 対向 電気部品 コネクタ カバー 筐体 突出 収容空間 制御回路 内面 電子制御ユニット 固定 電気接続箱 貫通孔 開口部 ケース内 蓋ハウジング ブラケット 収容部 挿入 収容 形成 一体 開口</p>	<h3>Z'37.部品・装置の配置</h3> <p>方向 配表 配置 モータユニット サイドメンバ 前方 車体 上面 左右 車両前方 鞍型 車両 上方 左右一対 フロアパネル 車両後方 後輪 パワーコントロールユニット 下方 バッテリーケース フレーム部材 車室 車幅方向 電池モジュール 搭載構造 車体 フレーム クロスメンバ 前後 上下方向 後方</p>	<h3>Z'38.パーツなどの移動、位置</h3> <p>本体 操作 作業 全閉状態 表示部 方向 閉塞 可動部材 所定位置 位置 脚部 アンロック位置 移動方向 係合 移動+できる ロック位置 可動パネル 付勢力 視認+できる 移動 係 押圧 解除 ロックレバー 阻止 操作+できる 移動 許容 回転 車両前方 規制</p>	<h3>Z'39.構造の形成・方位</h3> <p>平行 凹部 交差 交互 外周 加熱部 突起 内面 先端 内周面 径方向 軸方向 導電性材料面 周方向 外周面 突出方向 環状 対向隣接形成 周縁 外側 間隔 断面 接触 垂直 部分 端面</p>	<h3>Z'40.支持構造</h3> <p>変位一端 連結 昇降 固定 回転 コイルばね 保持部材 他端 回転+できる ベース部材 昇降+できる 移動 揺動+できる 係 車体 移動+できる 両端 支持部材 上下方向 支持部 平行 回転+できる ブラケット 規制 支持 直交 立設 先端 縮</p>

diff-PLSAでは、通常のPLSAのトピックに加え、運転者の補助、発電・蓄電、燃料電池、情報の取得・提供、車両用灯具、組成物の製造などを含む50個のトピックが抽出されました

Z'41.装置やユニットの構成

液体
適合車両用ピストン周囲導入 形態
チューブ相互接続シャフト回転+できる チャンバ
長手方向軸受容取り外し+できるプレート 移動+できる
同軸ハウジング 固定軸方向協働領域 電気機械
モジュール 配置 出口ガイド 車輪入口

Z'42.システム・方法の構成

調整監視 送信信号 ID 推進
無線測定値 応答 構成 伝送測定 データ通信 受信
センサ信号センサ プロセッサ 確立 エネルギー貯蔵装置
送信機 制御信号方法磁界 システム受信機
生成 命令 結合 感知

Z'43.その他方法

形式 熱機関
調整 駆動+できる 作動装置 トルク 車両用駆動
独立ドライブトレイン 連結+できる 歯車列 方法 ブレーキシステム
作動+できる 依存 トランスミッション 駆動部 パワートレイン
運転+できる 内燃機関 自動車 電気機械 電気システム
結合+できる 駆動装置 電気式 車輪 制御 システム

Z'44.組成物の製造方法(樹脂や電解液など)

式重量工程 Cr
質量一種 存在下塗布反応 混合 群ケイ素
化合物 組成物 製造方法 ニッケルアルミニウム
錫 乾燥 含有 製造材料 被覆分散式中混合物
銅 金属 MnTi

Z'45.機能性組成物・成形品(耐熱性や耐衝撃性等)

材料 携帯電話 金型 樹脂性
耐薬品性 電気部品用途 成形品 外觀 耐衝撃性
用途 熱可塑性樹脂 電気部品 ポリアリレンスルフィド 流動性
自動車部品 樹脂 機械的強度 耐ドラッキング性 電気特性
耐熱性 熱伝導性 溶融流動性 透明性 電気絶縁性
機械的特性 成形品 外觀 成形性 難燃性
電子

Z'46.製造の効率化(小型化や低コスト化など)

大型化
効率簡素化 作業性 確保 防水性
軽量化 対応+できる コンパクト 騒音 構造 コスト 実現+できる
小型化+できる 軽量 部品点数 小型化 低コスト
作業 信頼性 必要+ない 耐久性 小型 安価 製造コスト
省スペース 放熱性 高効率 削減 実現

Z'47.不具合の防止(損傷、感電、盗難など)

侵入 信頼性 衝撃 衝突
劣化 未然 位置ずれ 破壊 異音環境下 異常不具合
意図+ない 感電 確保+できる 漏電 静電気 振動 損傷
起因 過熱 ノイズ 誤動作 ショート 影響 過電流
故障 安全 誤検出 盗難

Z'48.その他

変形 静止
メインリレー 閉回路 圧電素子 路面
直列車両駆動用 安全装置 遮断動機 伸長 制御対象
バッテリー制御装置 監視 ハイブリッド トラクタ 空気圧
タイヤ 負 動作+できる 電池パック バイパス 電気接続
一側面 トレーラ バッテリー 電荷 並列 正

Z'49.タービン発電と船舶・飛行機への応用

日帰り旅行 運用 出力発電 燃料費ゼロ
既存 安価電気駆動 製造物全部 発電量 発電原価 宇宙到達費用
液体酸素圧縮駆動 横軸H軸車 太陽光加熱器 製造
改善 永遠 電気+液体空気+過熱蒸気 蒸気供給設備D
空気圧縮 容積圧縮仕事率 マッハ強い 容積 燃費
静翼 全動翼 船舶 重力発電 飛行機 反転
自動車

Z'50.重力発電の活用による地球温暖化防止

地球温暖化 燃料費ゼロ 液体金属 電気駆動 先送り
サンゴ 工場 電化 全盛 発電量 増大 水銀 既存 火力 原子力 発電
船舶 重力 加速度 加速 全廃 二酸化炭素 排気 ゼロ 金属球
安価 発電量 大気圧 同速度 同容積 仕事率 既存 世界
全面 電化 住宅 全盛 全地球 温暖化 既存 蒸気 タービン 発電
水 タービン 海水 温度 上昇 ゼロ 落差 重力 発電 蓄電池 駆動
自然現象 高速化 圧縮 空気 加速 垂直 下方 重力 加速度

※Z'49とZ'50は特定の出願人による重複した要約内容の特許から抽出された

※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 (関係の強い上位5つの単語を赤色で表示している)

【比較1】 differential PLSAは頻度の低い要素でトピックが構成される

通常のPLSAとdiff-PLSAで抽出された同様の意味を示すトピックを比較すると、diff-PLSAの方がより具体的な表現がトピックの上位語となり、細かい要素技術を抽出できています

「ブレーキ」に関するトピック

「非接触受電などの給電」に関するトピック

各トピックにおいて所属確率Pの高い上位7つの単語・係り受けと、それぞれの出現頻度n(出現文章数)

通常のPLSA

Z05

P(X Z)	n(X)	X: 単語	P(Y Z)	n(Y)	Y: 係り受け
7.3%	779	ブレーキ	3.2%	92	車両-ブレーキ
5.0%	1,384	作動	2.4%	33	ブレーキ液圧-発生
3.4%	5,188	モータ	2.2%	53	制動力-発生
2.9%	279	制動力	1.8%	49	ブレーキ-備える
2.9%	867	運転者	1.7%	32	操作量-応ずる
2.7%	1,117	車輪	1.6%	82	ブレーキ-提供
2.6%	1,005	操作	1.6%	59	電気信号-基づく

Z14

P(X Z)	n(X)	X: 単語	P(Y Z)	n(Y)	Y: 係り受け
9.7%	1,162	給電	4.3%	1,350	電力-供給
5.1%	3,629	電力	2.7%	115	給電-行う
3.6%	1,140	電源	2.5%	52	電力-受電
3.4%	329	給電装置	2.0%	28	給電-電力
2.4%	4,655	電気自動車	1.6%	124	電源-接続
2.4%	180	非接触	1.5%	20	駐車装置-変化
2.4%	1,052	外部	1.5%	34	非接触-受電

所属確率の高い単語は全体の出現頻度も高い

全体の出現頻度が低い単語でも所属確率が高い

diff-PLSA

Z08

P(X Z)	n(X)	X: 単語	P(Y Z)	n(Y)	Y: 係り受け
2.2%	112	マスタシリンダ	3.2%	32	基づく-発生
2.0%	73	ブレーキ液圧	2.6%	32	操作量-応ずる
1.6%	72	ブレーキ操作	2.6%	33	ブレーキ液圧-発生
1.6%	117	液圧	2.5%	53	制動力-発生
1.4%	779	ブレーキ	2.3%	49	ブレーキ-備える
1.4%	279	制動力	2.3%	49	ブレーキ-備える
1.3%	111	操作量	2.3%	92	車両-ブレーキ

Z21

P(X Z)	n(X)	X: 単語	P(Y Z)	n(Y)	Y: 係り受け
3.2%	180	非接触	3.0%	52	電力-受電
2.5%	78	送電コイル	2.7%	28	給電-電力
2.4%	160	受電	2.3%	35	受電-電力
2.4%	86	受電コイル	2.1%	25	電力-給電
2.2%	329	給電装置	1.8%	20	駐車装置-変化
2.0%	73	受電装置	1.8%	20	駐車装置-変化
1.8%	124	受電部	1.8%	20	非接触-受電

※各トピックにおいて所属確率の高い順に名詞・係り受けを並べたとき、累積確率が50%になるまでの名詞・係り受けの平均頻度を対象に統計検定(Welchのt検定)を実施したところ、通常のPLSAとdiff-PLSAで1%有意の違いが見られた

【比較2】 differential PLSAでのみ抽出されるトピックがある

自動車の付加価値を高める重要な技術など、diff-PLSAでのみ抽出されるトピックが複数あり、頻度の高い要素を中心に典型的なトピックを抽出する通常のPLSAでは拾いきれません

「運転者の操作補助」に関するトピック

各トピックにおいて所属確率Pの高い上位7つの単語・係り受けと、それぞれの出現頻度n(出現文章数)

diff-PLSA

Z'09

P(X Z)	n(X)	X: 単語	P(Y Z)	n(Y)	Y: 係り受け
1.4%	50	シフトレンジ	1.9%	23	自動的-行う
1.0%	38	パーキングレンジ	1.7%	38	操作-行う
0.9%	147	検出結果	1.6%	22	駆動-停止
0.7%	1,200	停止	1.6%	26	動作-行う
0.7%	365	解除	1.6%	40	要する-時間
0.6%	34	キースイッチ	1.4%	46	停止-状態
0.6%	3,960	検出	1.2%	26	ブレーキ-作動
...

【解釈】運転者の操作を補助したり自動停止などの運転アシストに関する技術

特許の原文例

車両停止時の駆動源の切り替えによる車両の動き出しの防止又は車両の動き出しを運転者に知らせることのできる電気自動車を提供すること。アクセルペダル及びブレーキペダルの踏み込みがなく、エンジンの作動中であり、シフト位置が走行レンジである車両停止時に、クリープ走行制御の駆動源がモータからエンジンに切り替わった際のトルク変動により車両が動き出した際には、パーキングブレーキによる制動を自動的に行い、警報器及び警告灯による運転者への注意喚起を行う。

「情報の取得と提供」に関するトピック

Z'29

P(X Z)	n(X)	X: 単語	P(Y Z)	n(Y)	Y: 係り受け
1.2%	122	ナビゲーション装置	2.1%	48	情報-送信
1.1%	809	情報	2.1%	42	情報-含む
1.0%	165	目的地	1.8%	68	情報-取得
1.0%	111	位置情報	1.5%	31	情報-受信
0.9%	786	取得	1.5%	35	示す-情報
0.9%	819	送信	1.5%	126	情報-基づく
0.8%	558	表示	1.4%	25	情報-用いる
...

【解釈】位置情報を取得してドライバーにナビ情報として提供するという技術

特許の原文例

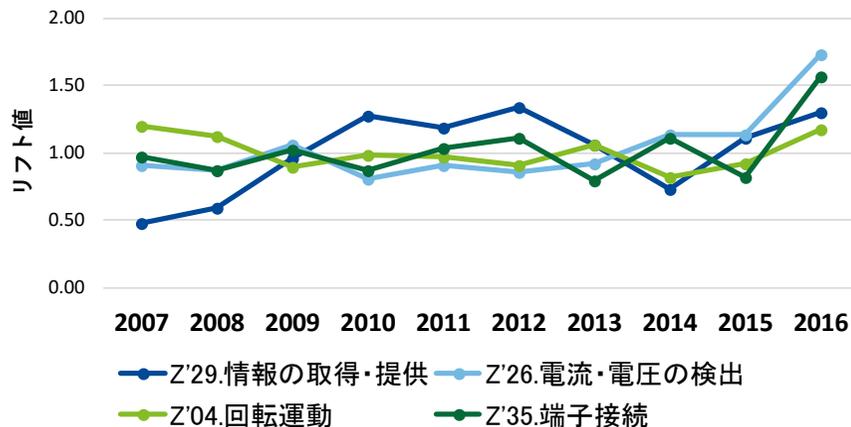
充電スタンドへの電気自動車の有効なナビゲーションを行う。本発明は、複数の充電スタンドの位置情報を記憶するデータベースと、電気自動車と通信して、電気自動車が搭載する電池の種別、エネルギー残量、エネルギー効率、現在位置を受信し、現在位置から電池残量およびエネルギー効率によって到達可能な充電スタンドを見つけ、その位置情報と、その充電スタンドが保持するエネルギー残量の情報を電気自動車に送信する充電スタンドロケータとを備える。

トピックをベースとした特徴の可視化

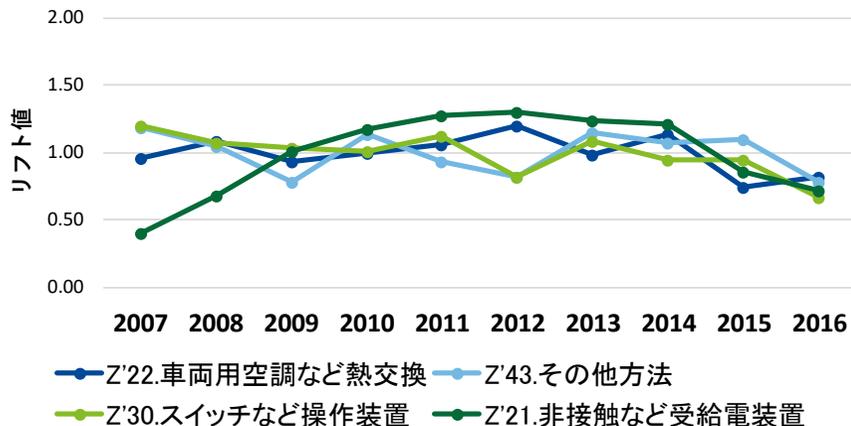
トピックのスコアを出願年や出願人の属性軸で分析することで、技術トレンドや競合他社のポジショニングを可視化でき、技術戦略を検討する上で効率的にインサイト獲得が期待できます

トピック×出願年のトレンド分析

近年上昇傾向にあるトピック



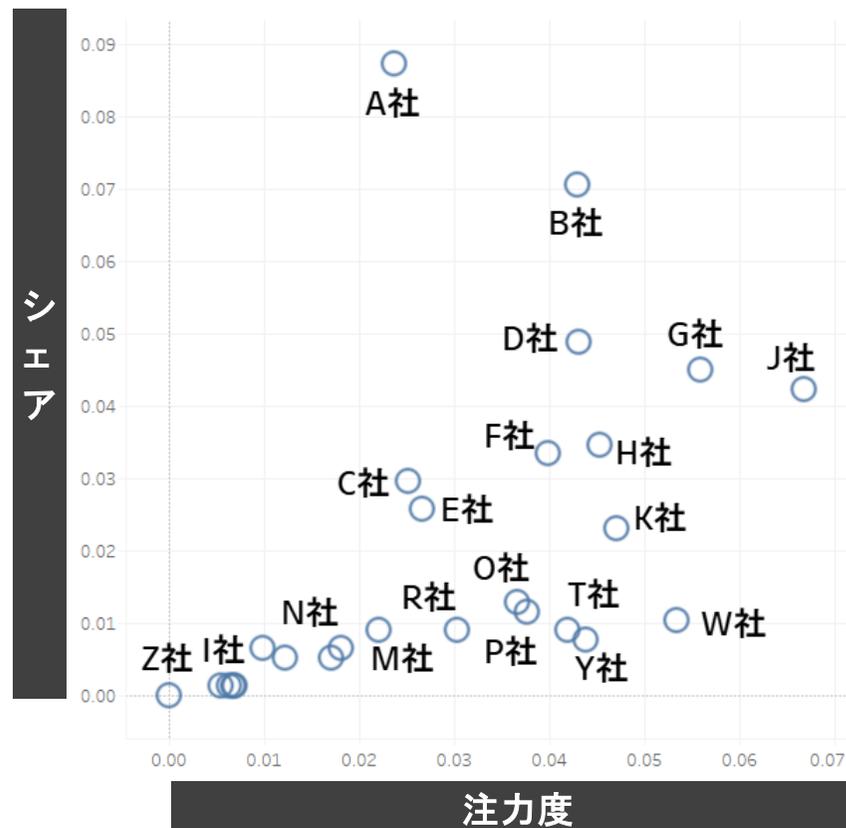
近年下降傾向にあるトピック



トピック×出願人の競合分析

各トピックに対する出願シェアと注力度を軸とした出願人のポジショニングマップ

「Z'29.情報の取得・提供」のポジショニングマップ



通常のPLSAでは全体の代表的なトピックを把握できますが、膨大で複雑なビッグデータから特に新たな気づきとなるインサイトを獲得したい場面では、*diff-PLSA*の適用は効果的です



全体を俯瞰する

通常のPLSA

- データ全体を表現する代表的なトピックを抽出できる
- 頻度の高い要素が優先的に反映される傾向がある
- 結果として典型的なトピックになってしまう

$P(Z)$
潜在クラス
 Z

$P(X|Z)$

X

行の要素

$P(Y|Z)$

Y

列の要素

インサイトを得る



differential PLSA



実測頻度の
共起行列 M

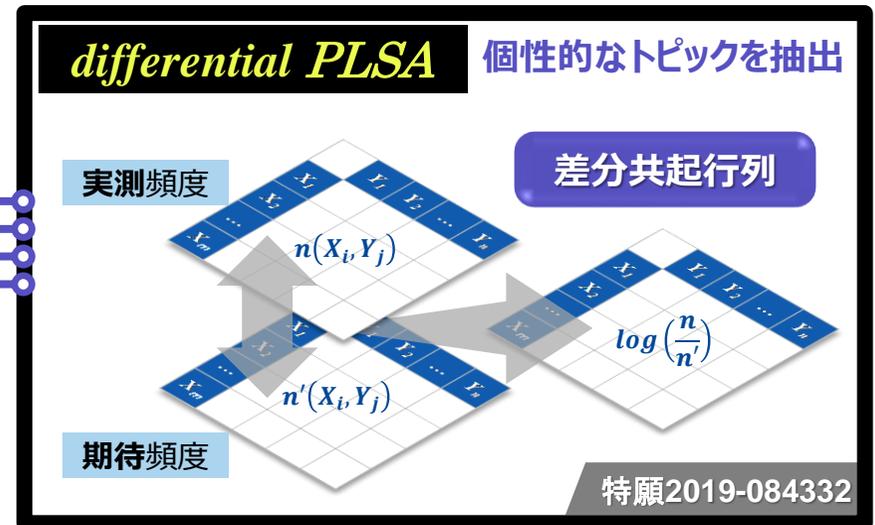
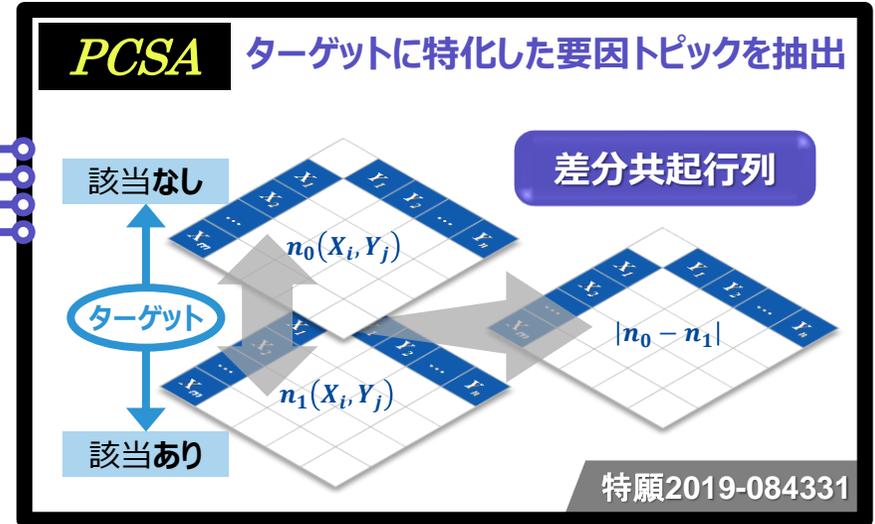
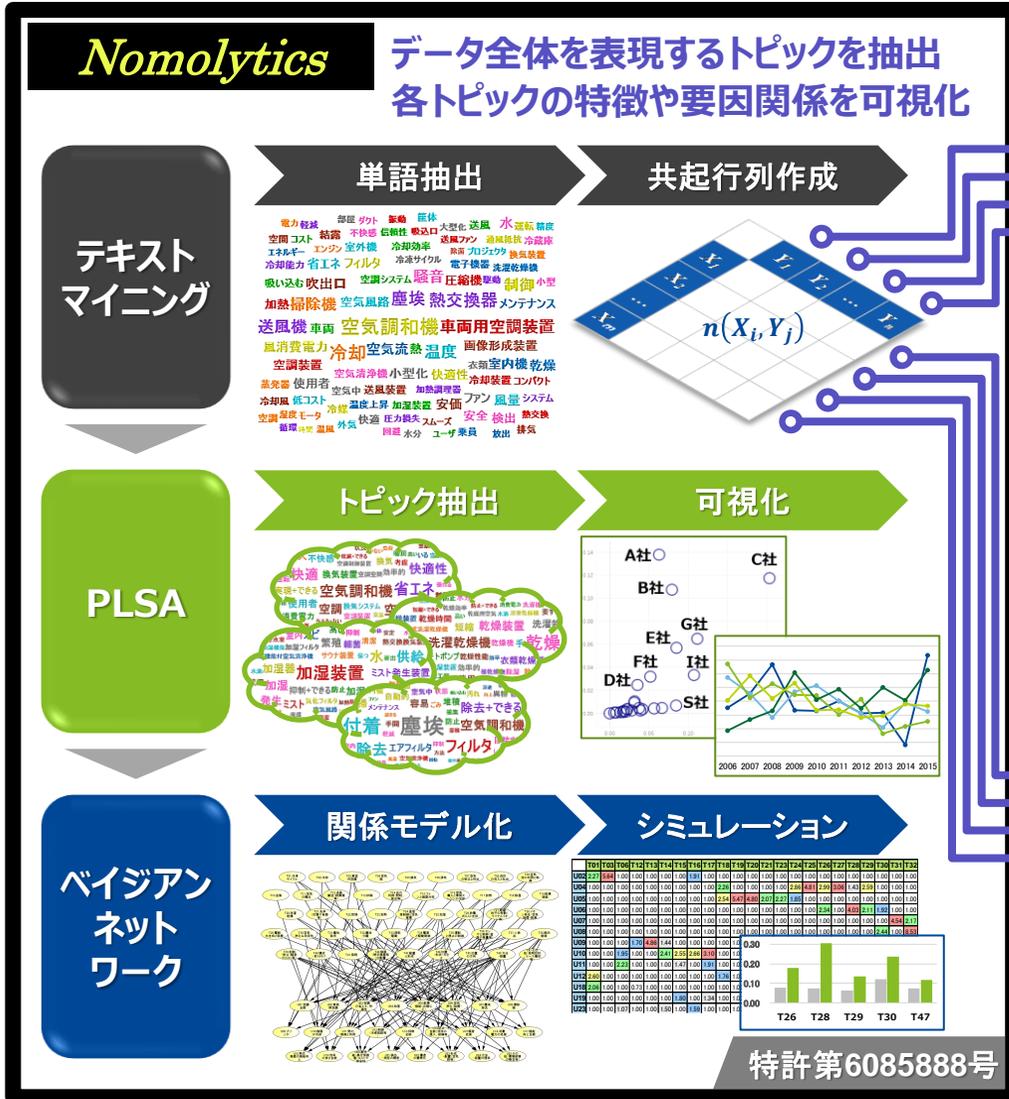
期待頻度の
共起行列 M'

- 頻度が低くても出現に特徴がある要素を引き立てられる
- より具体的で細かい要素で構成されるエッジの立ったトピックの抽出が期待できる
- ただし、結果はデータ全体を表現するものではない

テキストリファインニング技術のまとめ

テキストライニングについて開発した3つの技術

テキストライニング実行のアプローチとして、トピックモデルを中心とするAI技術を応用した
 ①Nomolytics、②PCSA、③differential PLSAという3つの技術を開発しました



Nomolytics、PCSA、diff-PLSA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事を把握
- 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
- トレ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

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

