日本計算機統計学会 第37回大会
「ビジネスでのデータサイエンス活用事例の紹介」

トピックモデルとベイジアンネットワークを応用したテキストデータの分析と活用

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

2023年6月3日

企業様のデータ分析・活用をご支援させて頂くコンサルティング会社で、これまでのアカデミックとビジネスの両方の経験を活かして2017年6月に設立しました

会社概要

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

Analytics Design Lab

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 ● 新しい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/

代表略歴：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
（専門領域：人間工学）
- 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

© 2023 Analytics Design Lab Inc.

テキストマイニング
×トピックモデル（PLSA）×ベイジアンネットワーク

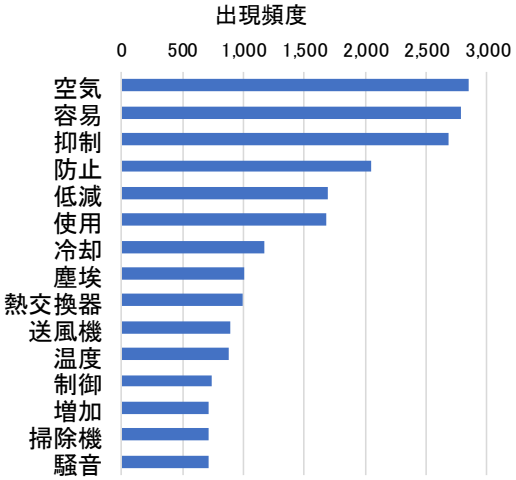
従来のテキストマイニングと課題

テキストマイニングは単語をベースに記述傾向の現状を把握しますが、結果が複雑で解釈が難しく、現状が変化したときに結果がどう影響受けるのかシミュレーションできません

テキストマイニングのアウトプット例

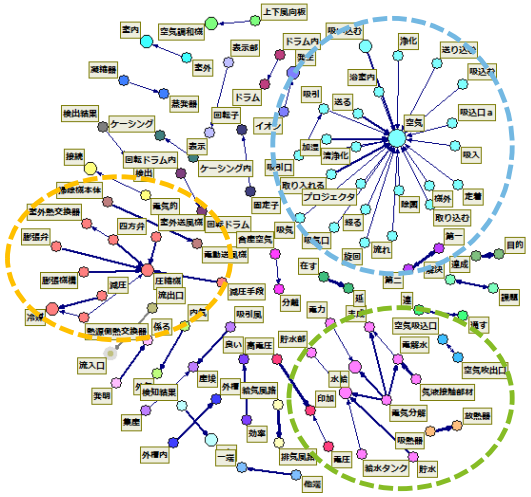
頻度集計

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



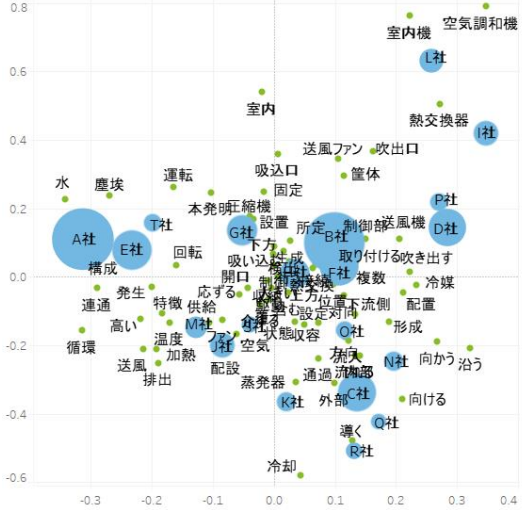
共起ネットワーク

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



コレスポンデンス分析

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

課題（1）

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
- ◆人が単語をグルーピングすることもあるが、その作業は主観的で負荷が大きい

課題（2）

現状把握に留まる

- ◆現状把握はできるが、その特徴が一般化（モデル化）されていない
- ◆現状から変化したときにそれに伴って結果がどう影響を受けるのか分からない

© 2023 Analytics Design Lab Inc.

複数のAI技術を応用したテキストデータの分析手法：Nomolytics®

【 特許第6085888号 】

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することで、ビジネスアクションに有用な特徴を把握可能にします

Nomolytics：Narrative Orchestration Modeling Analytics

テキストマイニング

文章に含まれる単語を抽出し、その出現頻度を集計する

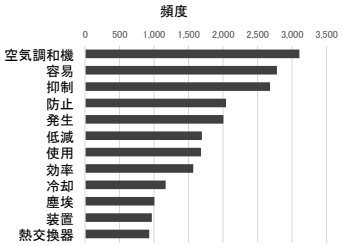
PLSA 確率的潜在意味解析

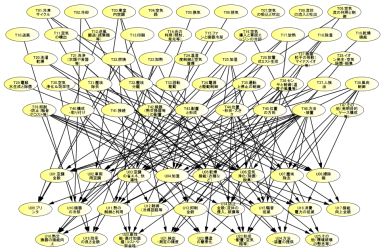
単語が出現する特徴を学習し、膨大な単語を複数のトピックにまとめる

ベイジアンネットワーク

トピックやその他属性情報など、テキスト情報内の要因関係をモデル化する

単語抽出 × *トピック類型化* × *要因関係分析*





Nomolyticsのメリット

膨大なテキストデータをいくつかのトピックという人間が理解しやすい形に整理し類型化できる

テキスト情報に潜む要因関係を構造化し、特徴を見たいターゲットのキードライバを発見できる

条件を変化させたときの効果を確率的にシミュレーションでき、有効なアクションを検討できる

© 2023 Analytics Design Lab Inc.

特許文書データの分析事例

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

データの抽出条件と抽出結果

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

分析データの加工

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

課題の文章

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

解決手段の文章

© 2023 Analytics Design Lab Inc.

Nomolyticsを適用した分析プロセスの全体像

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

トピックの抽出

特許の要約文に**テキストマイニング**を実行して単語や係り受け表現を抽出

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

PLSAの適用

【課題】の要約文から用途トピックを抽出

【解決手段】の要約文から技術トピックを抽出

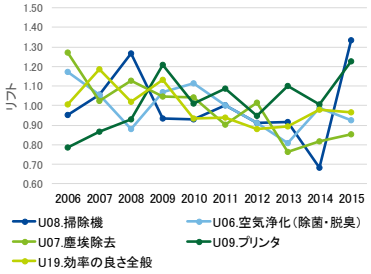
トピックのスコアリング

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	・・・	・・・	…	…	…	…	…	…

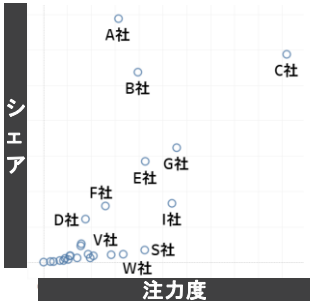
トレンドの分析

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



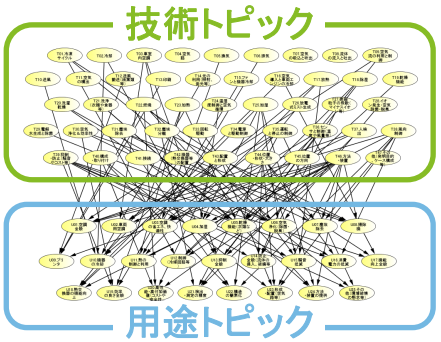
競合他社の分析

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動向を把握し、**事業化のための技術戦略**を検討する

技術⇒用途 の分析

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

© 2023 Analytics Design Lab Inc.

トピック抽出のアプローチ

テキストマイニングで単語と係り受け表現を抽出し、単語×係り受けの構成の共起行列に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 → 共起行列の作成 → PLSAの実行 → トピックの抽出

【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
・・・	・・・	…

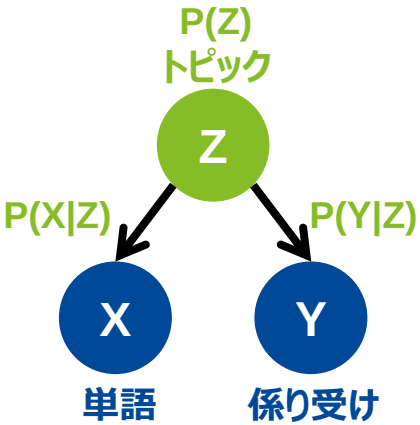
係り受け表現	頻度
空気調和機⇒提供	1,575
効率⇒良い	1,325
車両用空調装置⇒提供	578
掃除機-提供	545
容易-構成	539
画像形成装置-提供	334
抑制-提供	296
向上-図る	279
・・・	…

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

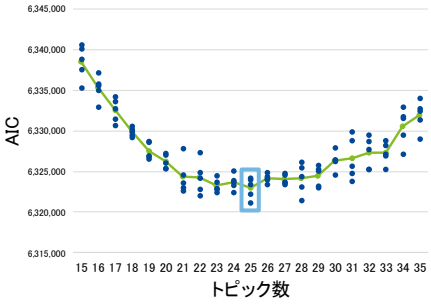
単語 ＼ 係り受け表現	空気調和機⇓提供	効率⇓良い	車両用空調装置⇓提供	掃除機⇓提供	…
空気調和機	1578	100	4	1	
空気	85	144	45	50	
容易	100	105	51	67	
抑制	142	95	64	63	
・・・					

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)

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 T32

P(Z) = 2.7%

P(X|Z)	単語
5.5%	送風機
5.2%	塵埃
4.1%	掃除機
3.6%	分離
3.5%	吸い込む
2.3%	集塵部
1.9%	配置
1.9%	集塵容器
1.6%	旋回
1.5%	含塵空気
…	・・・

P(Y|Z)	係り受け
2.1%	塵埃-分離
1.7%	分離-塵埃
1.7%	塵埃-含む
1.5%	吸い込む-塵埃
1.3%	含む-空気
1.0%	空気-分離
1.0%	送風機-吸い込む
1.0%	発生-送風機
0.9%	含塵空気-分離
0.9%	備える-掃除機
…	・・・

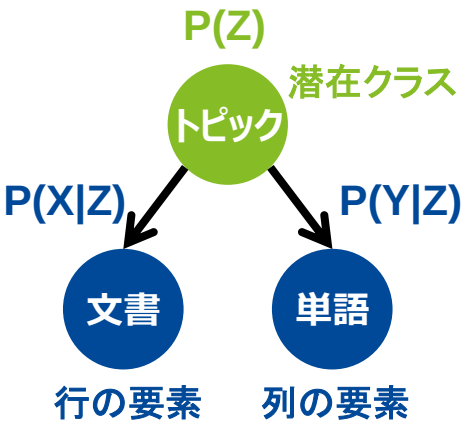
確率の高い構成要素から、トピックT32は「塵埃の分離」に関するトピックと解釈できる

© 2023 Analytics Design Lab Inc.

共起行列の構成の工夫

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

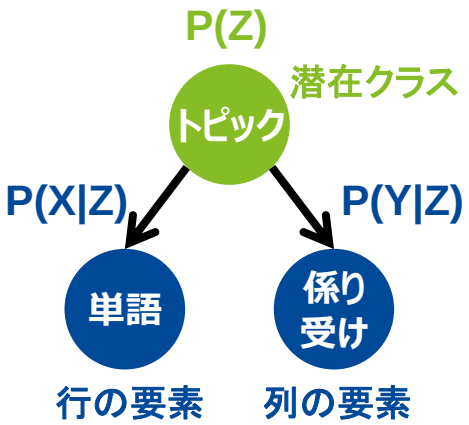
一般的なPLSAの共起行列



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

- 共起行列は"0"か"1"の2値で構成され、**ほとんどが"0"となる疎なデータであるため、要素間の違いが現れにくく、クリアなトピックを抽出しにくい**
- PLSAのトピックには行の要素と列の要素が同時に所属し、両方の情報軸からトピックの意味を解釈できるが、一方の軸（行）は文書IDという意味性の低い情報で、トピックの解釈に使用しにくい

NomolyticsでのPLSAの共起行列



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

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

© 2023 Analytics Design Lab Inc.

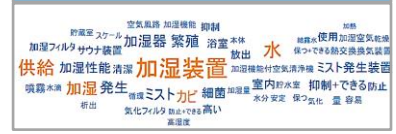
用途トピック25個の一覧

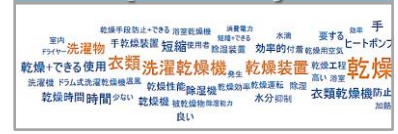
【課題】の文章からは、空調や加湿、空気清浄、掃除機、プリンタ、機器冷却、騒音や消費電力の低減、構造の簡素化などの用途が25個抽出されました

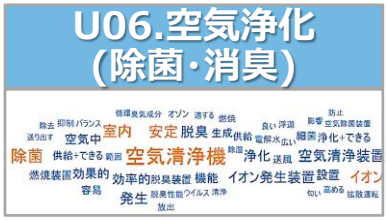
U01.空調全般

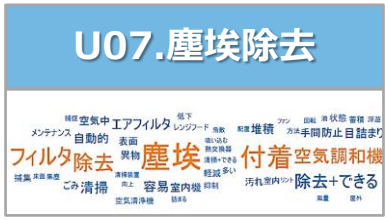
U02.車両用空調

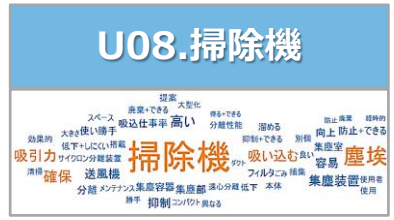
U03.空調の省エネ、快適性

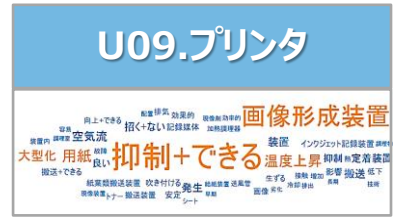
U04加湿


U05.乾燥機能(衣類など)


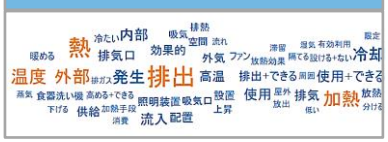
U06.空気浄化(除菌・消臭)


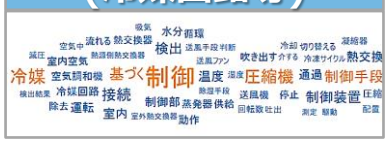
U07.塵埃除去


U08.掃除機


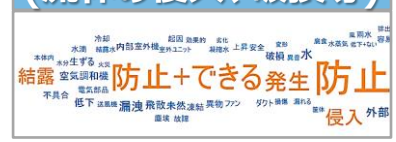
U09.プリンタ


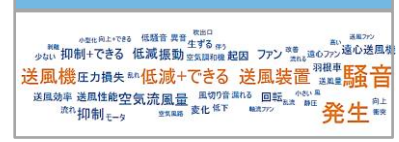
U10.機器の冷却

U11.熱の制御と利用


U12.制御(冷媒回路等)


U13.抑制全般

U14.防止全般(流体の侵入、破損等)


U15.騒音低減


U16.消費電力の低減

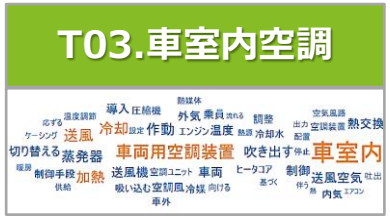
低減
消費電力
増加
維持
低減+できる
コスト
通風抵抗
製造コスト
空気抵抗
圧力損失
部品点数
ランニングコスト
乾燥機
乾燥時間
乾燥機能
乾燥機能
乾燥機
乾燥
乾燥
乾燥

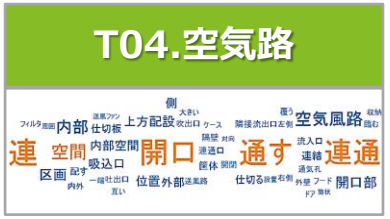
技術トピック47個の一覧①

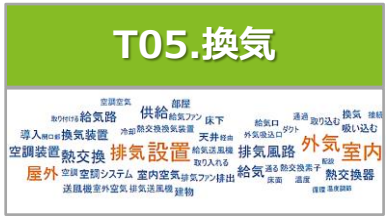
【解決手段】の文章からは、空気の冷却や換気、放熱、除湿、乾燥、加湿、イオン生成、空気清浄、塵埃分離、センサと制御、構成や配置などの技術が47個抽出されました

T01.冷凍サイクル

T02.冷却

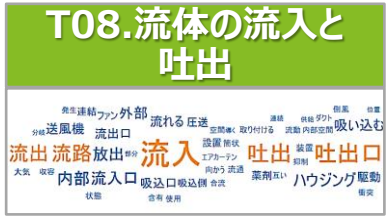
T03.車室内空調


T04.空気路


T05.換気


T06.排気

T07.空気の吸込と吹出

T08.流体の流入と吐出

吐出口
吐出
流入
流れる
流路
放出
流出口
流出
流路
流入
吸込口
吸い込む
内部流入口
吸込口
吸込側
含有
使用
ハウジング
駆動
葉剤
瓦い
装置
抑制
連結
供給
ダクト
エアカーテン
設置
衝撃
向かう
流通
状態
合流
流入
流れる
圧迫
空間
導く
取り付ける
流動
内部空間
流通
通路
流れる
圧迫
流出口
流路
放出
排出

技術トピック47個の一覧②

【解決手段】の文章からは、空気の冷却や換気、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

T26.放電式ミスト生成

T27.微細粒子の飛散(マイナスイオン等)

T28.イオン発生・空気除菌・脱臭

T29.電解水生成と除菌

T30.塵埃吸込＆効率性

T31.塵埃除去

T32.塵埃分離

T33.回転駆動

T34.電源と駆動制御

T35.運転と停止の制御

T36.センサと制御(温度や風量等)

T37.人検出

T38.風向制御

T39.抑制・防止(騒音やコスト等)

T40.構成・取り付け

T41.接続

T42.機器(熱交換等)の配置

T43.配置と形成

T44.位置・形状・大きさ

T45.位置の方向

T46.方法・装置

T47.その他(発明目的、ケース構成等)

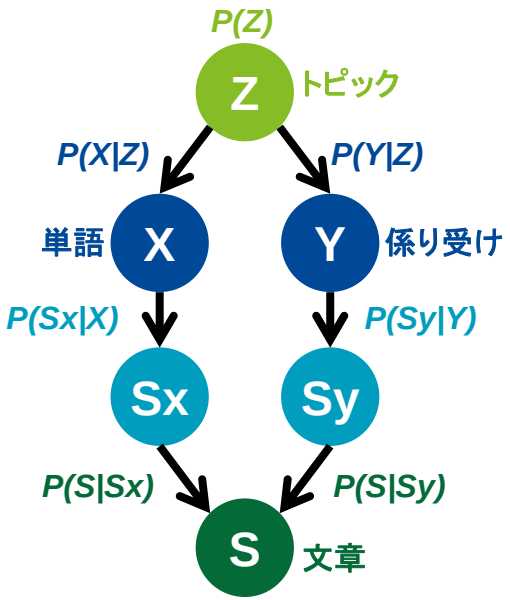
※文字の大きさはトピックに対する関係の強さを表現している
（上位5つの単語を赤色で表示している）

© 2023 Analytics Design Lab Inc.

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Sxと係り受けで定義される文章Sy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

$$Sx_h = \{X_1, X_2, \cdots, X_i\}$$

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

$$Sy_h = \{Y_1, Y_2, \cdots, Y_j\}$$

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
・・・					

© 2023 Analytics Design Lab Inc.

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、トピックを軸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た	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

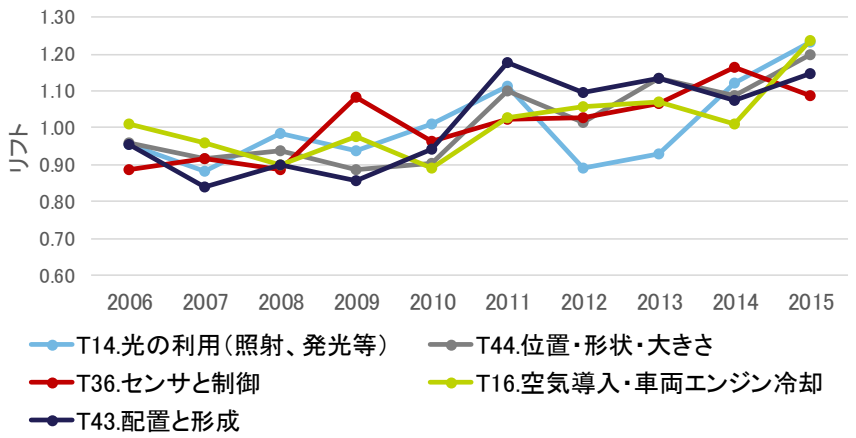
③用途と技術の関係分析

© 2023 Analytics Design Lab Inc.

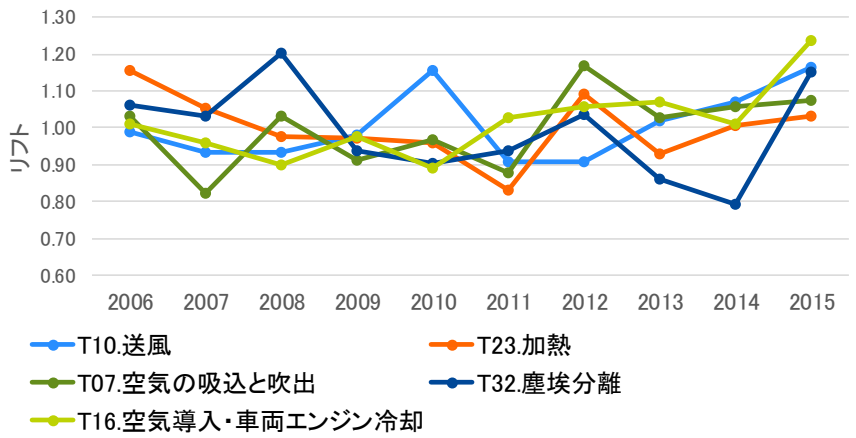
【①出願年の分析】 技術トピックの上昇トレンド

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タなどの光の利用に関する技術が上昇しています

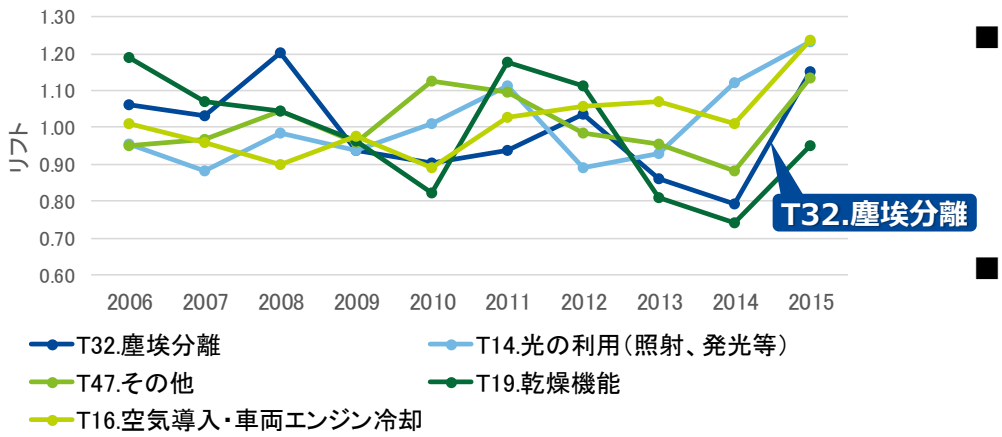
【長期】2006年からの上昇率 best5



【中期】2011年からの上昇率 best5



【短期】2013年からの上昇率 best5



集計の仕方

■ リフト値を出願年・トピックごとに集計

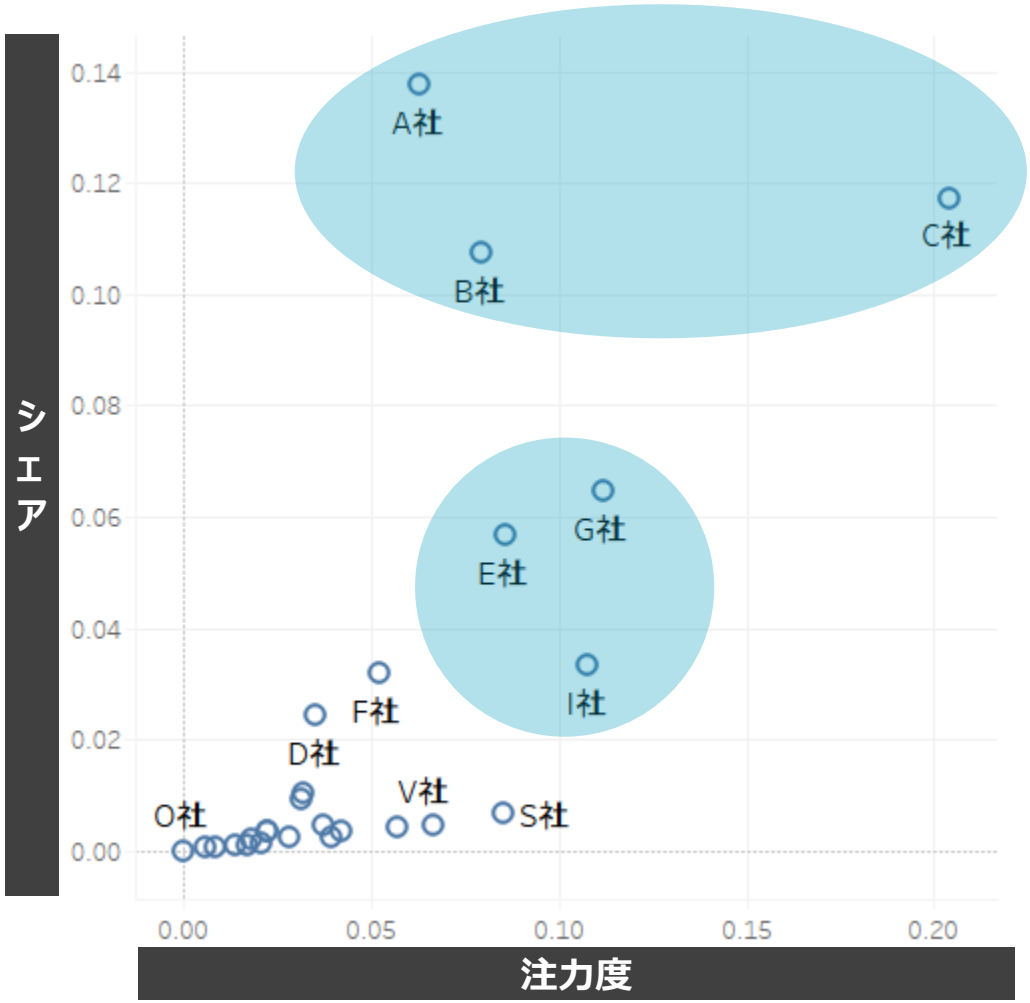
$$\frac{P(\text{出願年} \mid \text{トピックTx=1})}{P(\text{出願年})}$$

■ 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

© 2023 Analytics Design Lab Inc.

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

注力度とシェアの散布図



考察と戦略の検討

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

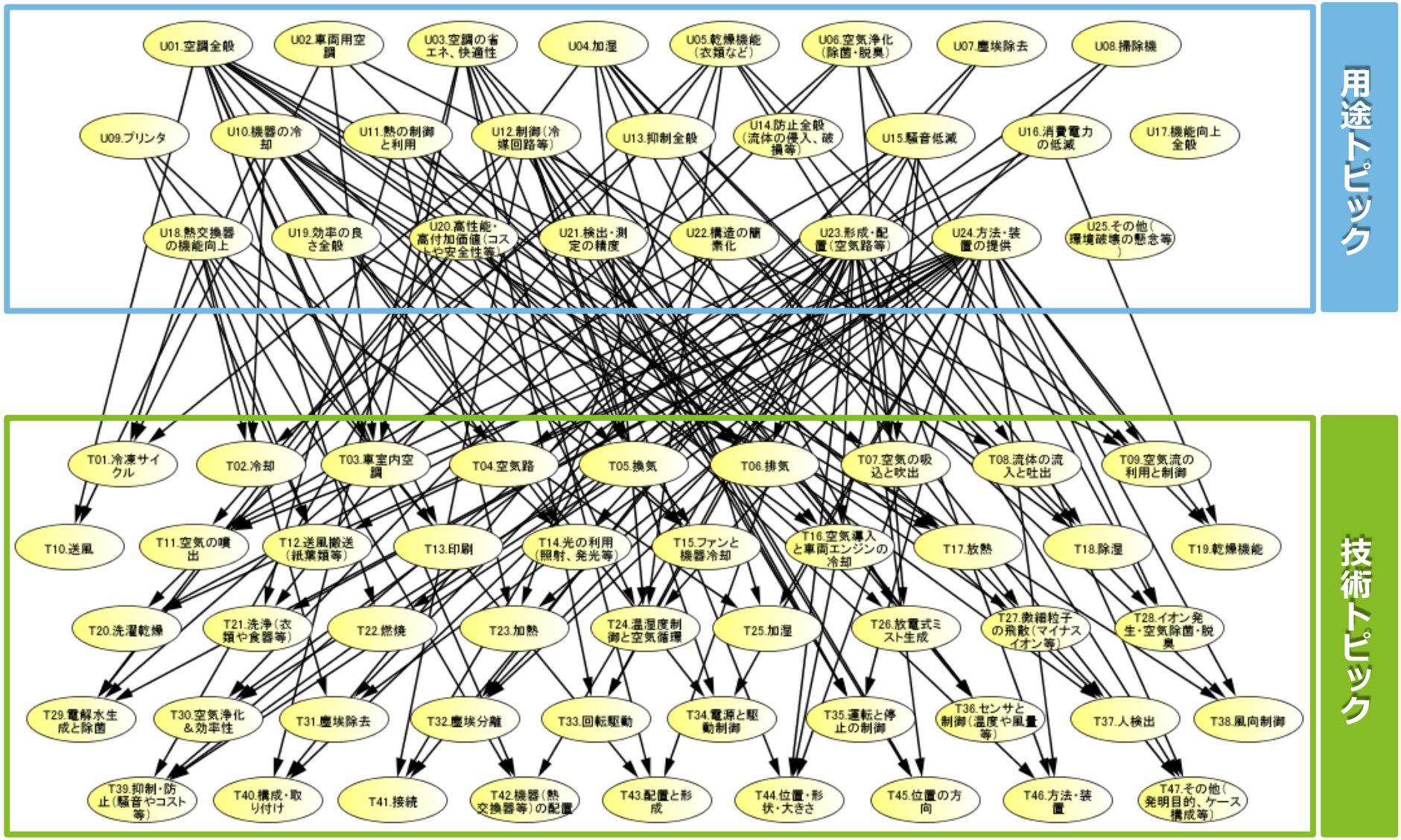
注力度とシェア

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

© 2023 Analytics Design Lab Inc.

【③用途と技術の関係分析１】 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

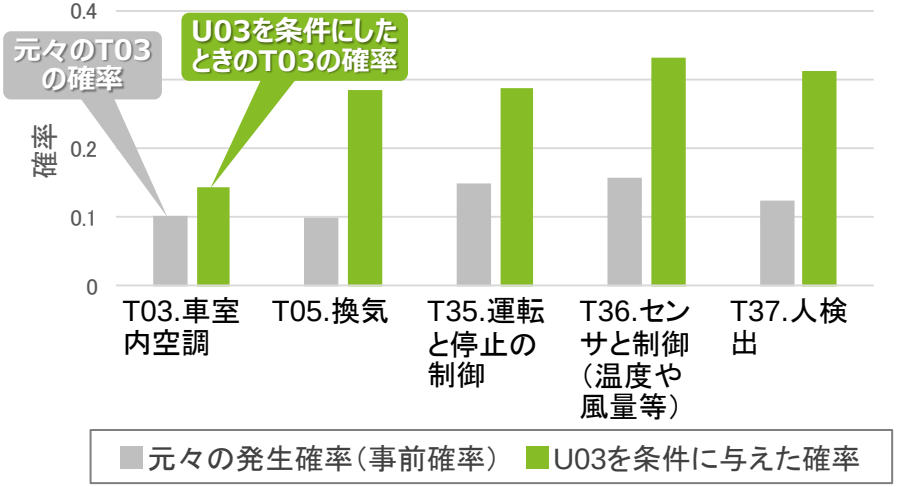


© 2023 Analytics Design Lab Inc.

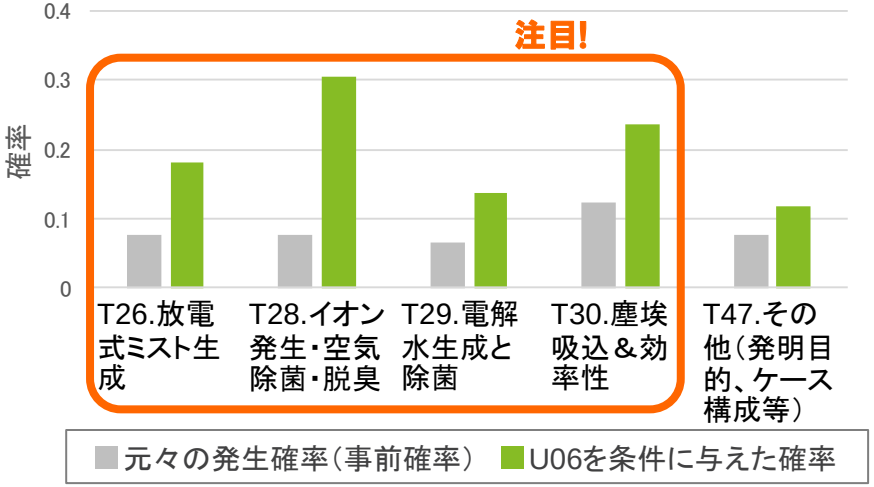
【③用途と技術の関係分析１】 用途と関係のある技術の確認

ベイジアンネットワークによって、1つの用途トピックを条件に与えたときの各技術トピックの確率の変化をシミュレーションし、その用途と関連の強い技術を確認します

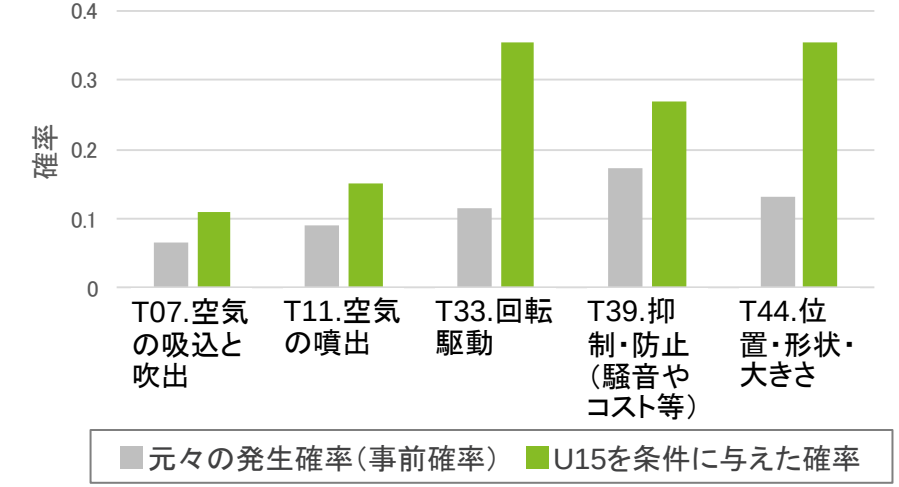
「U03.空調の省エネ、快適性」と関係のある技術



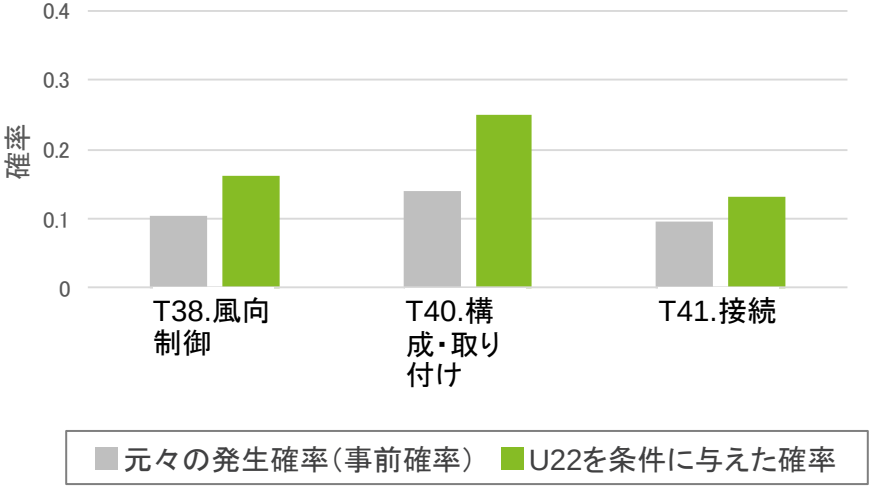
「U06.空気浄化」と関係のある技術



「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

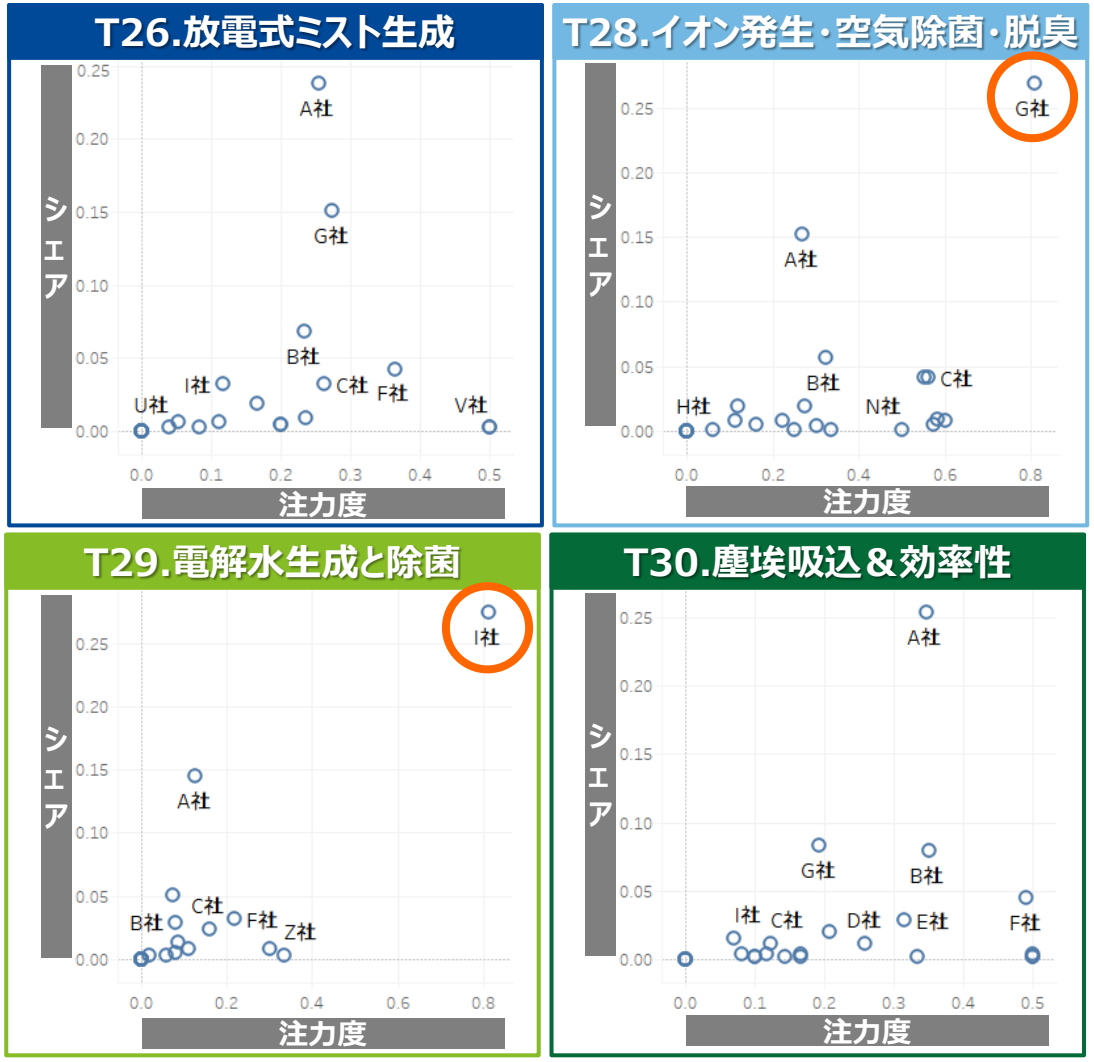


© 2023 Analytics Design Lab Inc.

【③用途と技術の関係分析１】 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



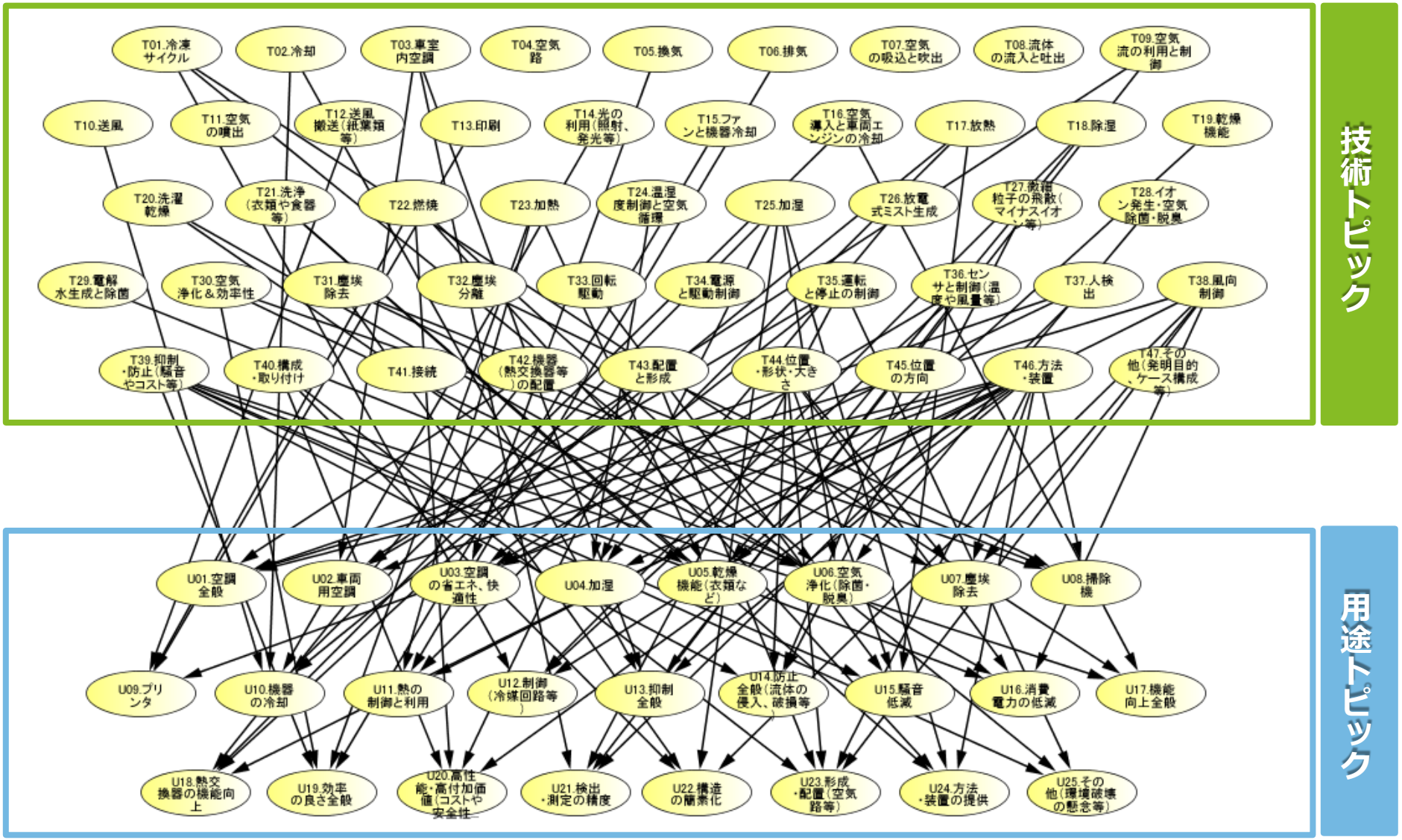
考察と戦略の検討

- 「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

© 2023 Analytics Design Lab Inc.

【③用途と技術の関係分析２】 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

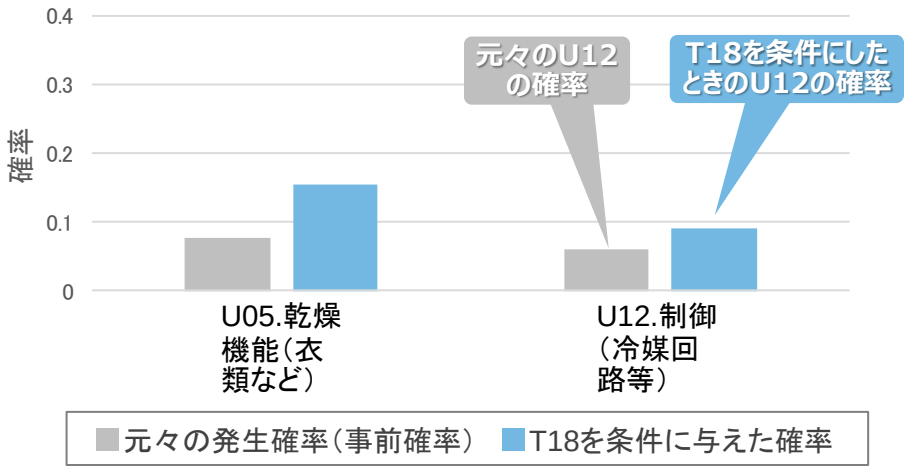


© 2023 Analytics Design Lab Inc.

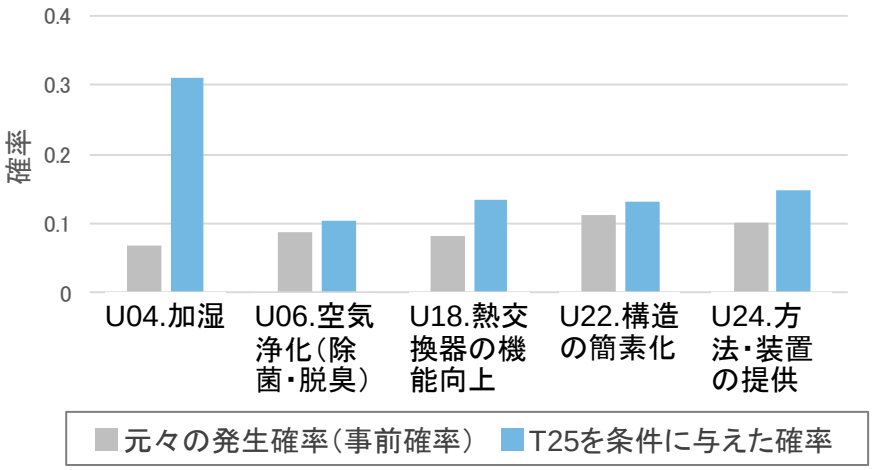
【③用途と技術の関係分析２】 技術と関係のある用途の確認

ベイジアンネットワークによって、1つの技術トピックを条件に与えたときの各用途トピックの確率の変化をシミュレーションし、その技術と関連の強い用途を確認します

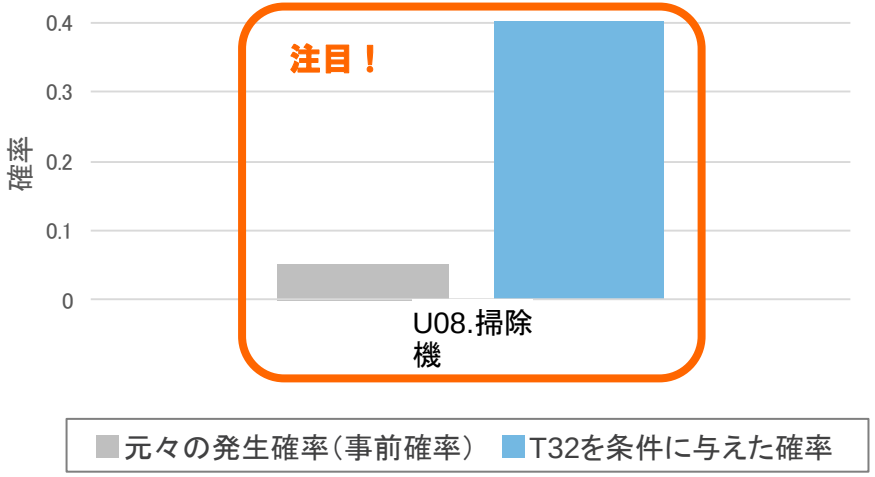
「T18.除湿」と関係のある用途



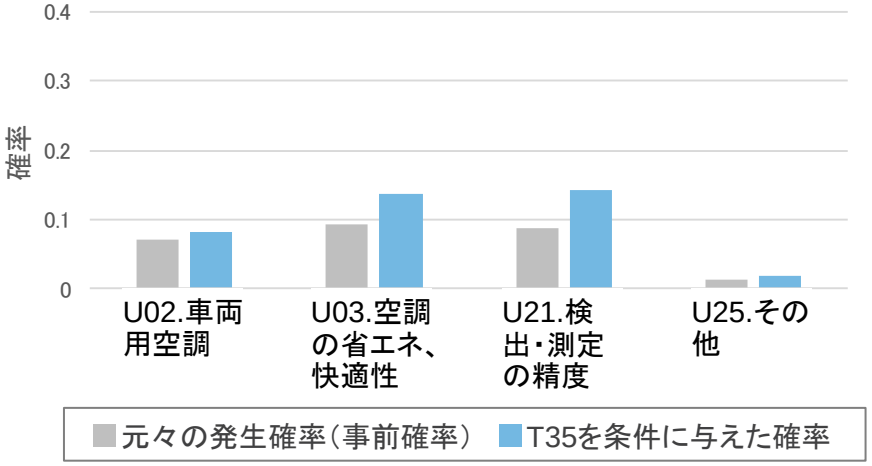
「T25.加湿」と関係のある用途



「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



© 2023 Analytics Design Lab Inc.

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

U08を想定したT32の特許例

発明の名称

電気掃除機

【課題】

集塵性能が向上しメンテナンスの軽減が図れる電気掃除機を提供すること。

【解決手段】

塵埃を含む空気を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。

U08を想定していないT32の特許例

発明の名称

画像形成装置

【課題】

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

【解決手段】

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

※対外説明用のため要約文は一部加工している

© 2023 Analytics Design Lab Inc.

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

サイクロン掃除機

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

サイクロン掃除機の技術はダイソンの様々な商品に応用されている

羽のない扇風機

空気清浄ファンヒーター

加湿器

ヘアドライヤー

ヘアスタイラー

ハンドドライヤー

© 2023 Analytics Design Lab Inc.

まとめ

複数のAI技術を応用したテキストデータの分析手法：Nomolytics®

【 特許第6085888号 】

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することで、ビジネスアクションに有用な特徴を把握可能にします

Nomolytics：Narrative Orchestration Modeling Analytics

テキストマイニング

文章に含まれる単語を抽出し、その出現頻度を集計する

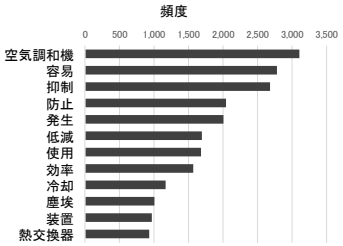
PLSA 確率的潜在意味解析

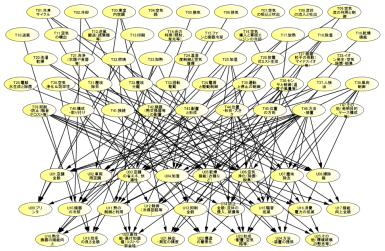
単語が出現する特徴を学習し、膨大な単語を複数のトピックにまとめる

ベイジアンネットワーク

トピックやその他属性情報など、テキスト情報内の要因関係をモデル化する

単語抽出 × *トピック類型化* × *要因関係分析*





Nomolyticsのメリット

膨大なテキストデータをいくつかのトピックという人間が理解しやすい形に整理し類型化できる

テキスト情報に潜む要因関係を構造化し、特徴を見たいターゲットのキードライバを発見できる

条件を変化させたときの効果を確率的にシミュレーションでき、有効なアクションを検討できる

© 2023 Analytics Design Lab Inc.

特許文書データへの適用

特許文書データにNomolyticsを適用することで、特許の要約をトピック化し、トレンドや出願人の競合分析をしたり、用途と技術の関係を分析することで技術戦略を検討できます

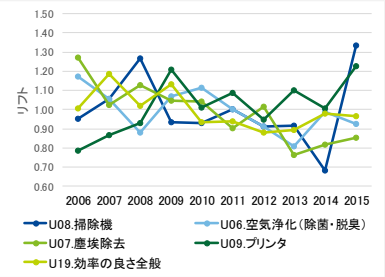
出願年・出願人×トピックの特徴分析

A 特許文書のトピック化

特許の要約にある【課題】と【解決手段】の文章にテキストマイニング×AIを適用することで、課題からは用途に関するトピックを、解決手段からは技術に関するトピックを機械的に抽出し、大量の特許の全体像を把握します

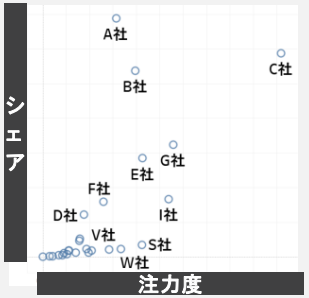
B トレンドの分析

出願年×トピックの関係を分析することで、用途や技術のトレンドを把握します



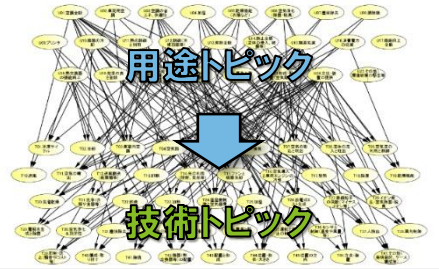
C 競合他社の分析

出願人×トピックの関係を分析することで、各社の特徴やポジションを把握します



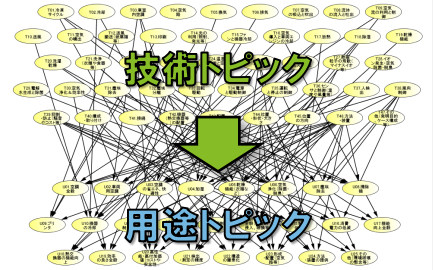
D 用途⇒技術の関係分析

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



E 技術⇒用途の関係分析

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



用途と技術の関係分析

© 2023 Analytics Design Lab Inc.

VOCデータへの適用（アンケートの自由記述、コールセンターの問い合わせ、Web口コミなど）

VOCデータにNomolyticsを適用することで、コメント内容をトピック化し、各属性の傾向や要因関係を分析することで、顧客目線の商品企画やマーケティング戦略を検討できます

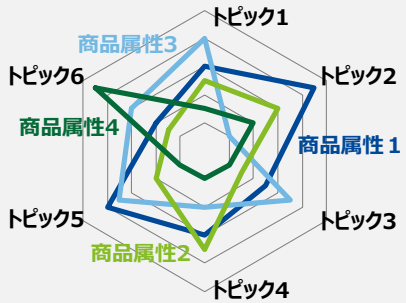
属性×トピックの1対1の関係を集計・可視化

A VOCのトピック化

VOCのテキストデータにテキストマイニング×AIを適用することで、コメントの内容を複数のトピックに機械的に類型化し、大量のコメントの全体像を把握します

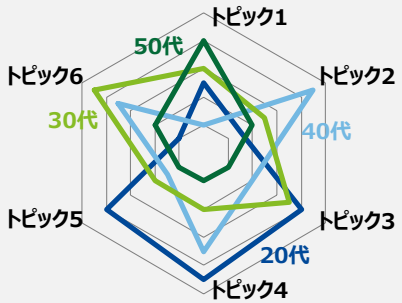
B 商品×トピックの傾向分析

商品やサービスの属性別に各トピックの関連度を計算することで、商品の各属性におけるコメントの特徴を把握します



C 顧客×トピックの傾向分析

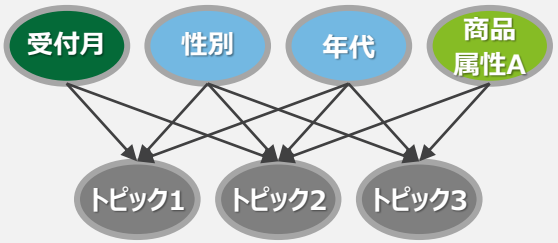
顧客の属性別に各トピックの関連度を計算することで、顧客の各属性がどのようなトピックに関心があるのか特徴を把握します



属性×トピックの複数間の関係構造をモデル化

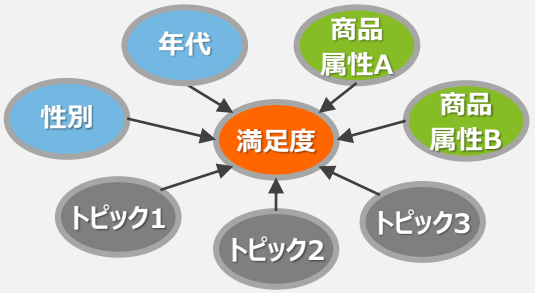
D トピックの要因関係の分析

各トピックに影響を与える要因を各属性情報（顧客属性、商品属性、時間属性など）からAIで探索し、そのトピックが発せられやすい要因条件を把握します



E 評価の要因関係の分析

データに満足度などの評価項目があれば、その評価に影響を与える要因をトピック及び各属性からAIで探索し、各要因条件からの評価予測や評価を高める条件を把握します



© 2023 Analytics Design Lab Inc.

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/

※ 資料の内容を引用または転載される場合は、
必ずその旨を明記いただくようにお願いします。

株式会社アナリティクスデザインラボ

Analytics Design Lab

© 2023 Analytics Design Lab Inc.