



確率的因果意味解析(PCSA)

テキストデータを用いたターゲット事象の要因トピックの抽出

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

2018年6月7日

会社紹介と自己紹介

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、というデータ分析を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

PLSAを用いた新しいテキストデータの分析手法の提案

PLSA(確率的潜在意味解析)とは

PLSAは、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です

PLSAの概要

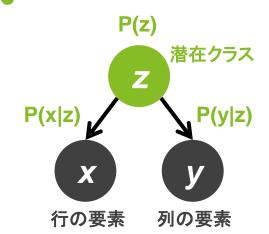
- 行列データの行の要素xと列の要素yの背後にある共通特徴となる潜在クラスzを抽出する手法である
- 元々は文書分類のための手法として開発されている (Hofman, 1999)
- 各文書の出現単語を記録した文書(行)×単語(列) という高次元(列数の多い)共起行列データに適用す ることで複数の潜在トピックを抽出し、文書(行)×ト ピック(列)という低次元データに変換して文書を分類 する

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

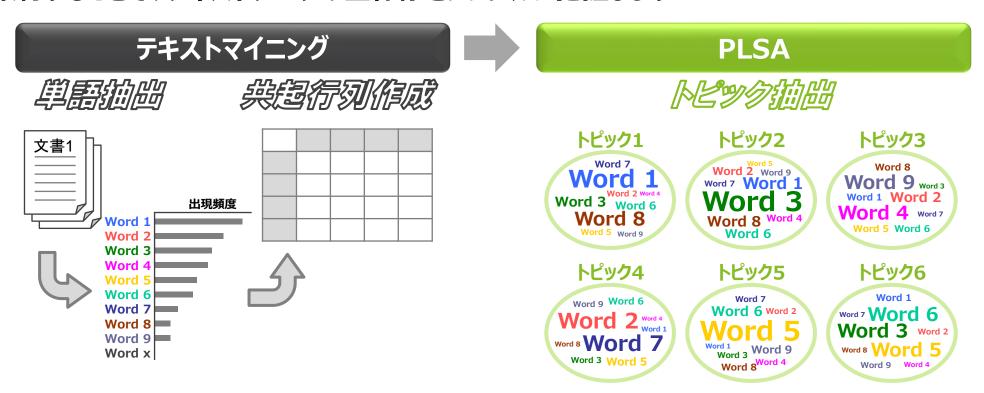
ソフトクラスタリングできる

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

4 2018年度 人工知能学会全国大会

【従来手法】PLSAを用いたトピックの抽出

テキストマイニングで抽出した単語で構成された共起行列にPLSAを適用して単語をトピックに 集約することで、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) PLSAで抽出したトピックを変数として扱い、他の属性情報との関係をベイジアンネットワークなどでモデリングすることで、その属性(ターゲット事象)に関連のあるトピックを探索できます

Nomolytics®: Narrative Orchestration Modeling Analytics

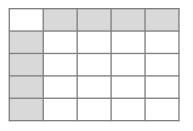
テキストマイニング

PLSA

ベイジアンネットワーク















トピック2



トピック4



トピック6

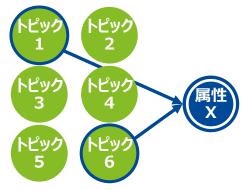


PEYDOZZZIL

	デー タID	属性 X	トピッ ク1	トピッ ク2	トピッ ク3	•••
	1	有	1	0	1	
	2	無	0	0	0	
	3	無	0	1	1	
Ì						

全データに対する各トピックの該当有無{0,1}を計算する

電景リング



トピックと他の属性情報(ターゲット事象)との関係をモデル化する

例えば、ロコミのコメントから得られたトピックとロコミの得点との関係をモデル化 することで、ロコミトピックの中から満足度向上に効果のあるものを探索できる PLSAで抽出したトピックを変数として扱い、他の属性情報との関係をベイジアンネットワーク などでモデリングすることで、その属性(ターゲット事象)に関連のあるトピックを探索できます

Nomolytics®: Narrative Orchestration Modeling Analytics

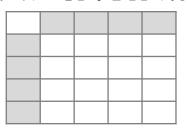
テキストマイニング

PLSA

ベイジアンネットワーク











例えば、口コミのコメントから得られたトピックと口コミの得点との関係をモデル化 することで、ロコミトピックの中から満足度向上に効果のあるものを探索できる

PLYDOZZENE

デー タID	属性 X	トピッ ク1	トピッ ク2	トピッ ク3	•••
1	有	1	0	1	
2	無	0	0	0	
3	無	0	1	1	

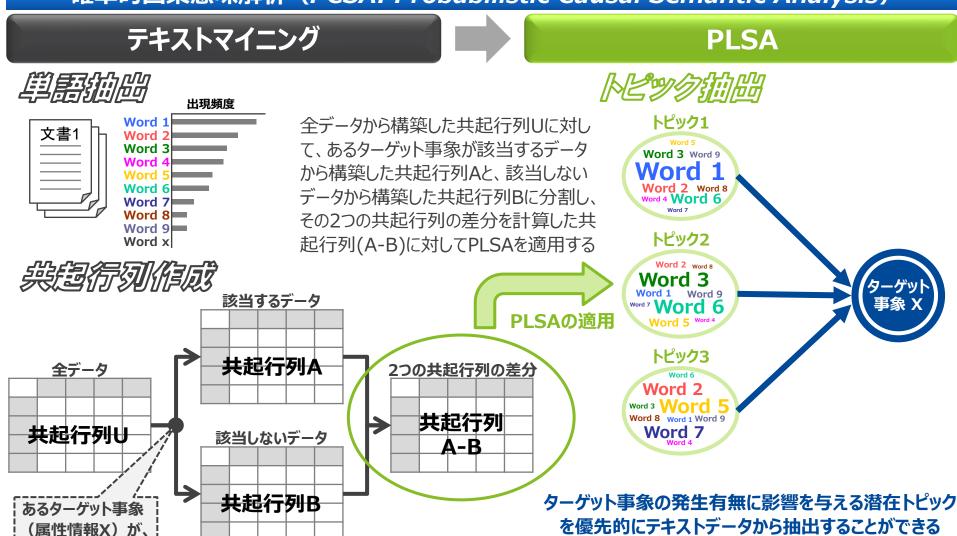
全データに対する各トピックの該 当有無{0,1}を計算する



トピックと他の属性情報(ターゲッ ト事象)との関係をモデル化する

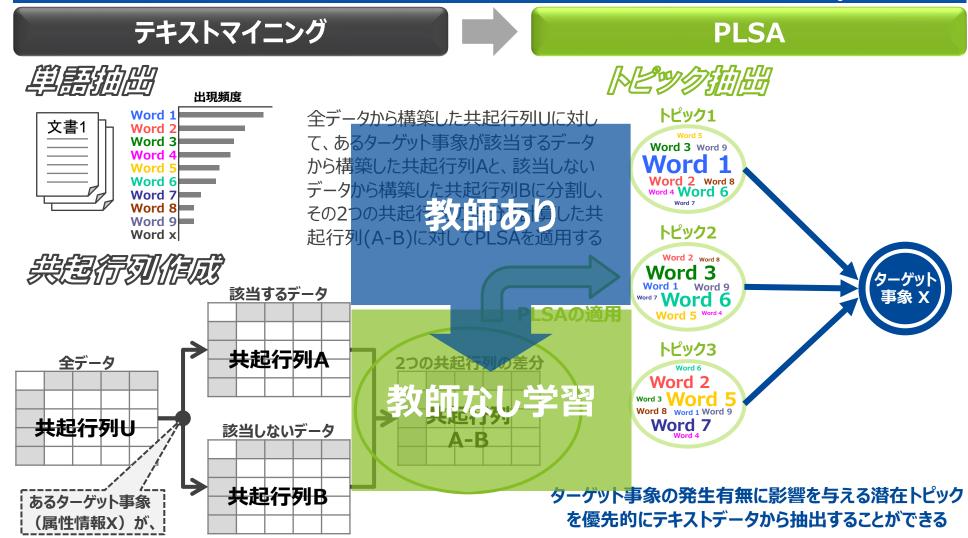
ターゲット事象の該当データと非該当データからそれぞれ構築した2つの共起行列の差分に PLSAを適用することで、ターゲット事象に影響を与えるトピックに限定して抽出します

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



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、ターゲット事象に影響を与えるトピックに限定して抽出します

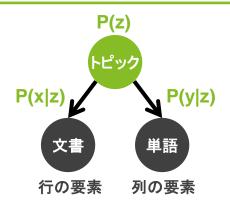
確率的因果意味解析(PCSA: Probabilistic Causal Semantic Analysis)



【補足】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×単語」とすることで、解釈のしやすいトピックが抽出でき、異なるデータの共起行列でも加法性が成り立ちます

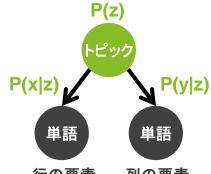
通常のPLSA



	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、データ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
- データが異なれば共起行列の構成も異なる

本研究で適用するPLSA



行の要素 列の要素

	単語a	単語b	単語c	単語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、 データ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
- 行も列も単語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
- 異なるデータの共起行列でも、行列の単語の構成が共通していれば共起行列の加法・減法が適用できる

分析事例

「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

出願年が

- ①2012年以前(22,387件)
- ②2013年以後 (7,652件)

①ターゲット事象との関係を考慮しないPLSAの実行

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、約3万件の特許の全体を構成している技術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

【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
配置	名詞	6,630
供給	名詞	3,692
内部	名詞	3,531
送風機	名詞	3,454
制御	形容詞	3,428
位置	名詞	3,185
吸い込む	名詞	3,099
吹出口	名詞	3,078

係り受け表現	頻度
空気⇒吸い込む	1,112
吸い込む⇒空気	889
連⇒通す	754
備える⇒構成	700
空気⇒吹き出す	689
吹出口⇒吹き出す	670
空気⇒供給	648
空気⇒送風	643
•••	

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

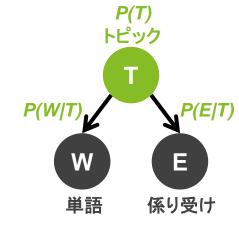
係り受け表現

		空気⇒吸い込む	吸い込む⇒空気	連⇒通す	備える⇒構成	• • •
	配置	210	191	150	145	
24	供給	127	101	67	108	
単語	内部	134	107	126	64	
12	送風機	225	197	112	95	

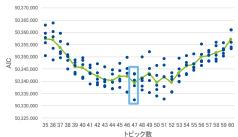
単語:5,187語×係り受け:7,174表現 (頻度10件以上を対象)

※「空気」「形成」「設ける」など、頻度が 突出して多いものはノイズになりやすい ため除外している

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(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

トピックt32

P(t32) = 2.7%

確率	単語	確率	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

確率の高い構成要素から、t32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

空気の冷却や空気路、換気、機器の冷却、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、人の検出、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

選及の万年表れる 関係弁 高発 法決決器 (1997年 新れ 近風 1888年 新北 法文 (1997年 新北 大 (1997年 1997年 新北 (1997年 1997年 1

t02.冷却

t03.車室内空調

t04.空気路

利 (本) 空気風路 高が (大きに アイル) (本) 空気風路 高が (大きに アイル) 原理 内部 空間 開 (本) 連接 (本) 空気風路 高が (東京 中) 変込口 (東京 中) (東京

t05.換気

t06.排気

t07.空気の吸込 と吹出

新りた。 また (1855年) 日本 (1855年

t08.流体の流入と 吐出

発生 連載 ファン外部 流れる 圧送 giánt この付ける 海内 原産 9年 ※20 送風機 流出口 流出 流路 放出 20 流出 流路 放出 20 元出 流路 放出 20 元の 20 元の

t09.空気流の利用と制 御

t10.送風

t11.空気の噴出

t12.送風搬送 (紙葉類等)

t13.印刷

t14.光の利用 (照射、発光等)

東端上方発光管 光 中心 無線 向かう進行方向 光 中心 機線 発光部 (株内 アレクタ 反 女 側 光源 反射 (水) 流通 及射線 冷却空気 (ネラブロシアウ 反 位置 ダクト光東 間口部 照射 (大) 新田校社 面配比 光源を置いる (大) 大地 東側口部 照射 (大) 大学 (大) 大

t15.ファンと機器冷却

t16.空気導入と 車両エンジンの冷却

t17.放熱

t18.除湿

t19.乾燥機能

海市の 日報 国 日報 接出 セールブ加加学品 日本 1752年 日本

t20.洗濯乾燥

すが送風 通風器 泉い込U 保用光準 排出 内部 工程様気 モータ 回転動 乾燥工程 水槽 秋楽 駆動 乾燥回転自在 兵総 乾燥排 牧容 回転ドラム温風 洗濯物 木洗濯乾燥機 加熱手段 送風手段 太類支持 保風服 乾燥度 循環 東交き付き風歌 展記フンで値体 内側 開水 配置 建 英風影 加熱 回転

t21.洗浄 (衣類や食器等)

t22.燃焼

t23.加熱

t24.温湿度制御と ____空気循環

t25.加湿

15 2018年度 人工知能学会全国大会

※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)

空気の冷却や空気路、換気、機器の冷却、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、人の検出、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.空気清浄& 効率性

共内の経験後、別所、受引力 202口 使用者 変化検出 浄化 作用面域 大きい 3点を設 発生四点を連 支別権送機性供できる 良し 1 空気節 イライナーできる 単空気清浄機 得る 1年 応する 空気中 得る 1年 応する 203 年 203 年

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

属て8 世内の 松 周囲配列 第5 伝熱管 直交 貫通内部 並べる流通 延びる 第5 伝熱管 直交 貫通内部 並べる流通 延びる 第7 風上側フイン 間隔 長手方向 互い配置 総部積層 方向 所定熱交換器 ※成本の 上のでは があります。 表面書名 現ま板状 (入) のよう を組みせる いまからなった。

t44.位置・形状・大きさ

対応 側し 長き離れる部分突出方向回廊 現なき延びるベルマウス 対向 小さい 軸方向 位置 中央部 回転軸 大きい向かう 見る配度外周側アーシップ しー・形状 羽鹿畑 領域 沿う外部 傾斜 側 回転方向 増部 プレード 形状 羽鹿畑 領域 滑り 同一 手紙 東側 同一 日本 ・ 大阪中心 ・ 大阪中の ・ 大阪 ・ 大阪中の ・ 大阪 ・ 大阪中の ・ 大阪中の ・ 大阪中の ・ 大阪 ・ 大阪中の ・ 大阪中の ・ 大阪中の ・ 大阪中の ・ 大阪中の ・ 大阪中の ・ 大阪の ・ 大阪の

t45.位置の方向

t46.方法•装置

表面 風車 第3,144年 編集 編集 編集 編集 編集 編集 編集 編集 2万人 高い工程配置 設置 (東京中) 新原五力 治却 必要生ずる 発電機 提供 受ける 方法 移動 風力発電装置 水 生成 使用 発電 回転ステップ作る 発症 装置 発生 タッー 流れ 発音 数差 発生 タッー 流れ

t47.その他(発明目的、 ケース構成等)

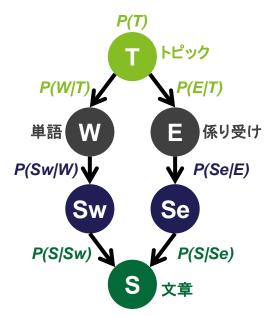
各データに対するトピックの該当有無の計算

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した

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

 $Sw_h = \{W_1, W_2, \cdots, W_i\}$

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_i で定義される文章 Se_h

 $Se_h = \{E_1, E_2, \cdots, E_J\}$

トピック T_k を条件とした文章 Se_h の出現確率

 $P(Se_h|T_k) = \sum_{i} P(Se_h|E_j)P(E_j|T_k)$

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_i の出現文章数の逆数)

 $P(Se_h|E_i) = 1/n(E_i)$

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Sh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		1.6

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		1.6
•••					

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		0
•••					

トピックと出願年の関係

トピックと出願年の関係を示す指標値(リフト値)を①2012年以前②2013年以後で計算し、2013年前後での増減率を集計したところ、高いものから低いものまでばらついています

リフト値の増減率	2012年 以前の リフト値	2013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
12.9%	0.97	1.09	t14.光の利用(照射、発光等)
12.2%	0.97	1.09	t45.位置の方向
11.3%	0.97	1.08	t10.送風
10.6%	0.97	1.08	t03.車室内空調
9.6%	0.98	1.07	t37.人検出
8.9%	0.98	1.07	t41.接続
7.9%	0.98	1.06	t40.構成・取り付け
7.4%	0.98	1.05	t07.空気の吸込と吹出
6.5%	0.98	1.05	t33.回転駆動
6.3%	0.98	1.05	t34.電源と駆動制御
3.9%	0.99	1.03	t04.空気路
3.7%	0.99	1.03	t17.放熱
3.3%	0.99	1.02	t01.冷凍サイクル
3.1%	0.99	1.02	t08.流体の流入と吐出
2.5%	0.99	1.02	t25.加湿
1.9%	1.00	1.01	t11.空気の噴出
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

リフト値の増減率	2012年 以前の リフト値	2013年 以後の リフト値	トピック
-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
-2.2%	1.01	0.98	t02.冷却
-2.3%	1.01	0.98	t47.その他(発明目的、ケース構成等)
-3.8%	1.01	0.97	t21.洗浄(衣類や食器等)
-4.4%	1.01	0.97	t13.印刷
-4.8%	1.01	0.96	t05.換気
-5.0%	1.01	0.96	t15.ファンと機器冷却
-6.4%	1.02	0.95	t12.送風搬送(紙葉類等)
-7.3%	1.02	0.94	t29.電解水生成と除菌
-9.8%	1.03	0.93	t32.塵埃分離
-9.9%	1.03	0.92	t27.微細粒子の飛散(マイナスイオン等)
-15.5%	1.04	0.88	t24.温湿度制御と空気循環
-15.7%	1.04	0.88	t30.空気浄化&効率性
-17.9%	1.05	0.86	t20.洗濯乾燥
-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

トピックと出願年の関係を示すリフト値

※出願年を①2012年以前、②2013年以後として計算

P(出願年Y | トピックT=1)

P(出願年Y)

②ターゲット事象との関係を考慮したPCSAの実行

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

2つの共起行列の作成

差分の共起行列の作成

PLSAの実行

トピックの抽出

①出願年が2012年以前のデータ22,387件(文章数:33,283件)

		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	
出	供給	95	65	53	81	
単語	内部	93	64	92	44	
пы	送風機	145	134	79	76	

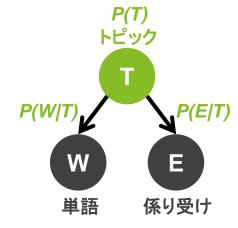
②出願年が**2013年以後**のデータ 7.652件(文章数:11,831件)

			係りき	とけま	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	
単	供給	32	36	14	27	
菩語	内部	41	43	34	20	
HE	送風機	80	63	33	19	

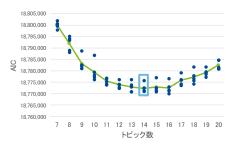
①2012年までの共起行列と ②2013年からの共起行列の 差の絶対値を計算した共起行列 を作成する

		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	
22	供給	1.8	12.9	4.8	1.8	
芦吾	内部	7.9	20.3	1.3	4.4	
=1	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(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

トピックT10

P(T10) = 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、T10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

T01.冷凍サイクル

温度圧縮 送風 切り替える 車室内 吐出熱 数交換器冷媒回路冷却 膨張弁 接続 熱交換器 /王 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

T02.空気の冷却

通す 淳入室内 屋外熱交換連 外部 配設<mark>内部冷却外気</mark>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<mark>排気</mark> 吹き出す 吹出口 循環設置 開口 供給 連通

T03.冷却ファン

無吸い込む 取り込む 加熱 載置
 外気筐体内排気口光収納 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 ピートシンク 冷却風 内部送風位置 収容

T04.空気流 (吸込と吹出)

T05.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 加熱送風部定着装置トナー像 シート 制御部 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

T06.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 <mark>衣類</mark> 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

T07.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良い 燃焼ガス バーナ 混合 燃料放出熱 ノズル 乾燥

T08.イオン発生

付着放電高電圧 電圧 生成 保持 吹出口 送風 オゾン帯電電極 発生 対向電極 分解送風路 空気中空気風路 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

T09.電解水の生成

室内 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 回転浸清

T10.塵埃の分離

T11.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 回転車 ロータ 取り付ける 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

T12.検出と制御

測定動作 第出 決定 人 駆動 検出 記憶温度センサ 温度 運転 判定 検出 物御手段 基づく制御部 制御 印転数 制御装置設定 風量 停止 応ずる 所定 室内出力 援態 取得

T13.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

T14.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 取り付ける
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

トピックと出願年の関係

トピックと出願年の関係を示す指標値(リフト値)を①2012年以前②2013年以後で計算し、2013年前後での増減率を集計したところ、増減率の高いものと低いものに集中しています

リフト値の増減率	يلا	012年 l前の フト値	2013年 以後の リフト値	トピック
48.5%		0.89	1.32	T13.車両用空調の配置
30.3%		0.93	1.21	T14.配置と形成
23.9%		0.94	1.17	T04.空気流(吸込と吹出)
17.7%		0.96	1.13	T12.検出と制御
14.9%		0.96	1.11	T11.羽の回転
13.1%		0.97	1.09	T03.冷却ファン
11.6%		0.97	1.08	T02.空気の冷却
7.0%		0.98	1.05	T01.冷凍サイクル
-0.3%		1.00	1.00	T05.紙葉類の搬送
-5.9%		1.02	0.96	T07.空気の燃焼
-8.8%		1.02	0.93	T09.電解水の生成
-17.0%		1.05	0.87	T10.塵埃の分離
-24.2%		1.07	0.81	T06.衣類乾燥
-28.7%		1.08	0.77	T08.イオン発生

※リフト値の増減率の高い順に並べている

全体の共起行列で抽出したトピックでは47個あるうち、 増減率は高いものから低いものまでばらついていた

			Par 0000 C100 3 2 0 C0 7 C
リフト値の増減率	2012年 以前の リフト値	2013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

トピックと出願年の関係を 示すリフト値

※出願年を①2012年以前、②2013年以後として計算

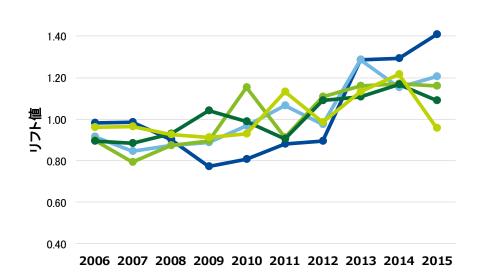
P(出願年Y | トピックT=1)

P(出願年Y)

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



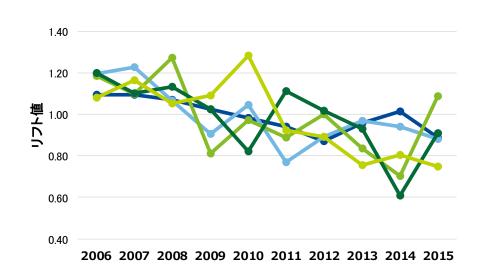
2013年前後の増減率

T13.車両用空調の配置48.5%T14.配置と形成30.3%T04.空気流(吸込と吹出)23.9%

T12.検出と制御 **17.7%**

T11.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

T07.空気の燃焼 -5.9% T09.電解水の生成 -8.8%

T10.塵埃の分離 **-17.0**%

T06.衣類乾燥 -24.2%

T08.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

「T13.車両用空調の配置」の特許要約例

発明の名称	出願年
車両用空調装置	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。

「T08.イオン発生」の特許要約例

発明の名称	出願年
扇風機	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

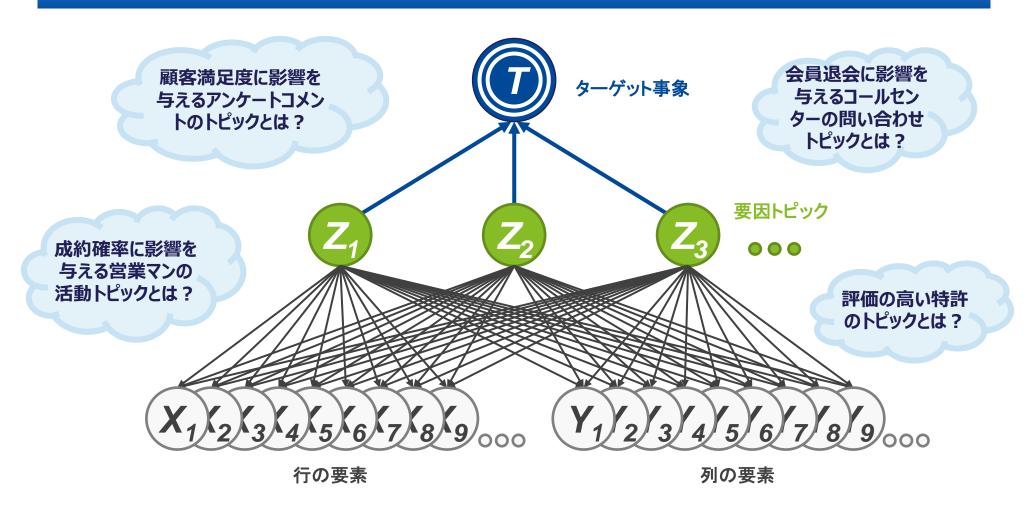
放電電極に水分を供給する貯水タンクおよび吸水管と を備え、放電電極と対向電極との間に高電圧を印加す ることで放電電極に保持される水分を霧化させる静電 霧化装置を具備する。したがって、人体に有益で、かつ 消臭や除菌等に効果があるイオンが水に包まれること で、寿命の長いイオンミストとして空中に放出することが でき、また水に包まれている分重いイオンミストを、扇風 機によって広範囲に拡散することができる

※対外説明用のため要約文は一部加工している

まとめ

PCSA(確率的因果意味解析)は、ターゲット事象に影響を与える要因トピックを優先的にテキストデータから自動構成でき、効果的なビジネスアクションの検討を支援できます

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



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

