Analytics Design Lab

ビッグデータ活用展2019 spring
株式会社NTTデータ数理システム　ブース内セミナー

ビッグデータのテキストマイニングに欠かせないAI技術

トピックモデルの適用と新たな応用技術

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5月8日、10日

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

株式会社アナリティクスデザインラボ

企業におけるデータ活用を支援するコンサルティング会社です。

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

ビッグデータのテキストマイニングに効果を発揮するAI技術

トピックモデルとは

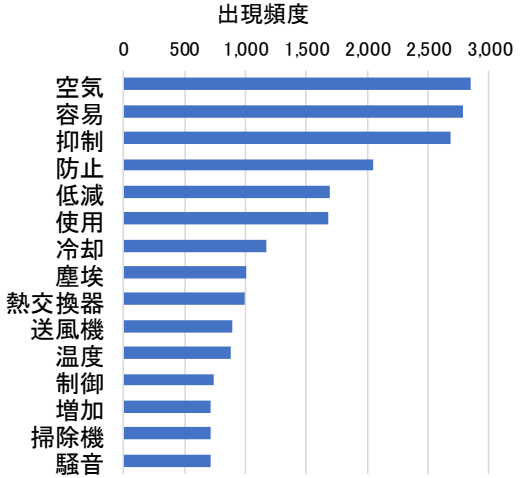
従来のテキストマイニングと課題

従来のテキストマイニングは単語をベースに全体像を把握しますが、膨大な単語で可視化される結果は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となります

従来のテキストマイニングのアウトプットの例

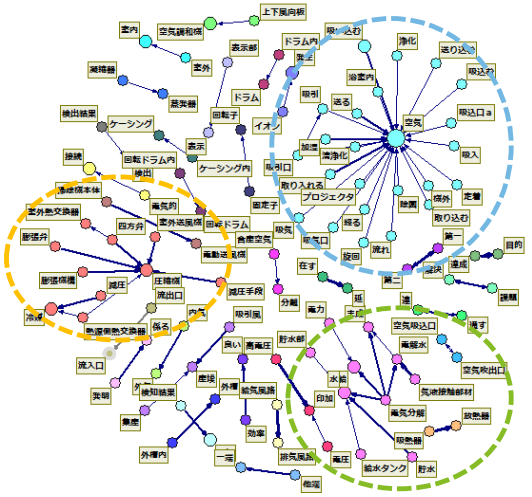
頻度集計

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



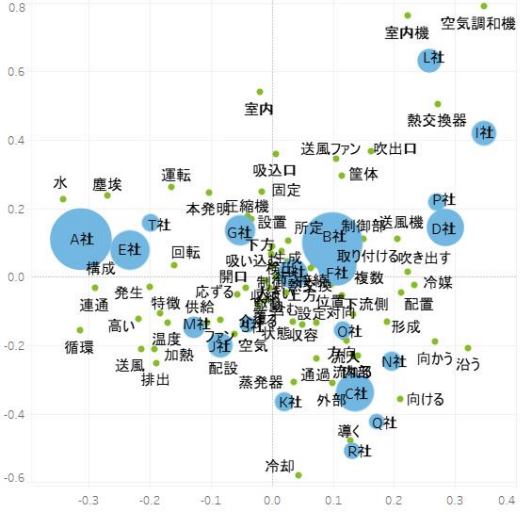
共起ネットワーク

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



従来の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

列数の多い複雑なデータのクラスタリングにはトピックモデルと呼ばれるAI技術が有効です

階層型クラスター分析

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

非階層型クラスター分析

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

LSA (Latent Semantic Analysis)

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

PLSA (Probabilistic Latent Semantic Analysis)

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

LDA (Latent Dirichlet Allocation)

- PLSAの拡張手法
- PLSA（他左3つの手法も含め）の過学習の問題に対して、LDAではディレクレ分布を仮定し新規データのクラスを推定できる
- 新規データに対応するため、抽出されるクラスは観測データを忠実に再現するものではなく、クラスの抽象度が高い傾向がある

従来のクラスター分析

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

トピックモデル

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

PLSA（確率的潜在意味解析）

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、複雑な観測情報でも分かりやすく理解できます

PLSAの概要

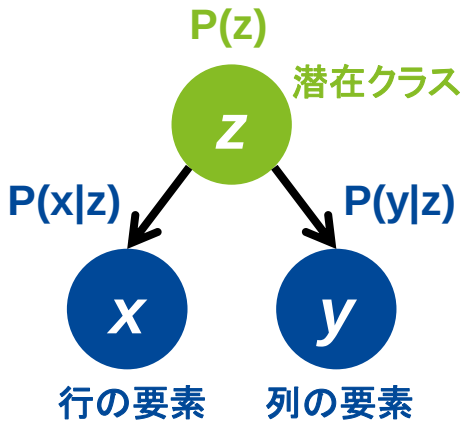
- 行列データの行の要素xと列の要素yの背後にある共通特徴となる潜在クラスzを抽出する手法である
- 元々は文書分類のための手法として開発されている(Hofman, 1999)
- 各文書の出現単語を記録した文書（行）×単語（列）という高次元（列数の多い）共起行列データに適用することで複数の潜在トピックを抽出し、文書（行）×トピック（列）という低次元データに変換して文書を分類する

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

PLSAの適用事例

電気自動車関連の特許文書データの分析

「車」「電気」を含む10年分の特許データ26,419件の要約文を対象に、テキストマイニングとPLSAでトピックを抽出し、各トピックの特徴を可視化します

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
- 要約文のテキスト情報

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を抽出

トピック化（PLSA） — Visual Mining Studio

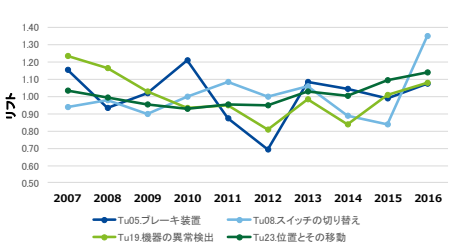
単語や係り受けを複数のトピックに集約

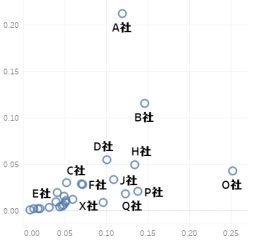
トピックのスコアリング

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属性を軸に集計・可視化





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 → 共起行列の作成 → PLSAの実行 → トピックの抽出

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
…	…

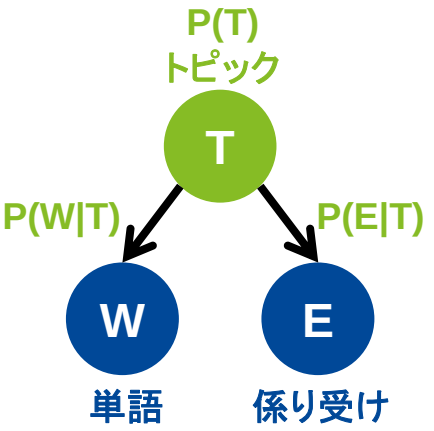
係り受け表現	頻度
電力⇒供給	1,208
否⇒判定	517
モータ⇒駆動	460
バッテリ⇒充電	440
効率⇒良い	419
供給⇒電力	332
電気自動車⇒提供	285
充電⇒行う	273
…	…

「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

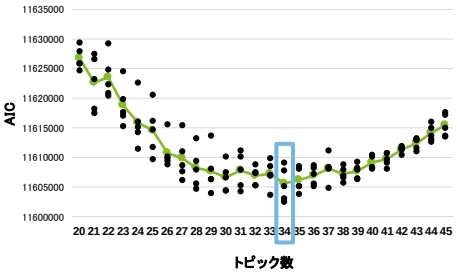
単語 \ 係り受け表現	電力⇓供給	否⇓判定	モータ⇓駆動	バッテリ⇓充電	…
構成	118	33	36	33	
制御	268	73	108	85	
配置	69	2	29	8	
モータ	239	61	494	58	
…					

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(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 Tu13
P(T) = 5.0%

P(W|T)	単語
12.6%	充電
8.9%	電気自動車
6.5%	蓄電装置
3.0%	バッテリ
2.0%	充電システム
2.0%	蓄電池
1.9%	電力
1.7%	制御
1.5%	充電スタンド
1.5%	放電
…	…

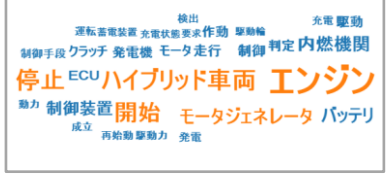
P(E|T)	係り受け
5.1%	バッテリ-充電
4.0%	充電-行う
3.9%	電気自動車-充電
1.9%	蓄電池-充電
1.6%	蓄電装置-充電
1.6%	電力-供給
1.3%	充電-開始
1.2%	電気自動車-接続
1.2%	充電-蓄電装置
1.1%	用いる-充電
…	…

確率の高い構成要素から、Tu13のトピックは「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

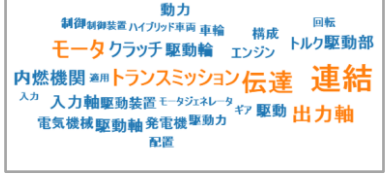
特許トピック34個の一覧①

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

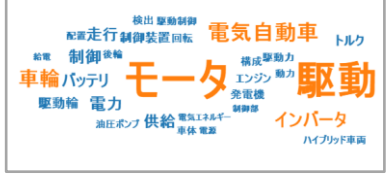
Tu01.エンジンの始動と停止



Tu02.動力の伝達

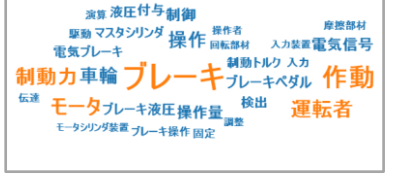


Tu03.モータ駆動



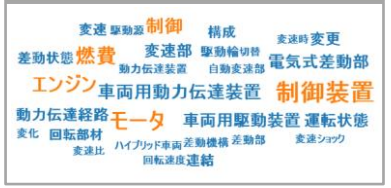
Tu04.ロータ・ステータなど回転部品の構成

Tu05.ブレーキ装置

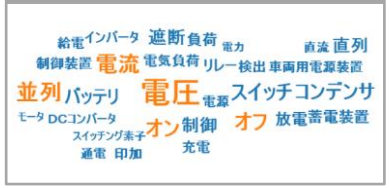


Tu06.動作制御

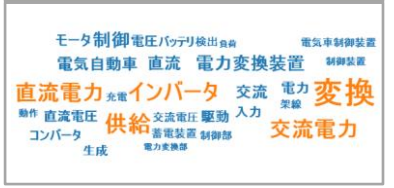
Tu07.動力伝達の制御



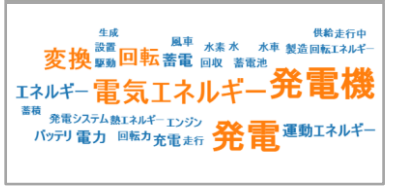
Tu08.スイッチの切り替え



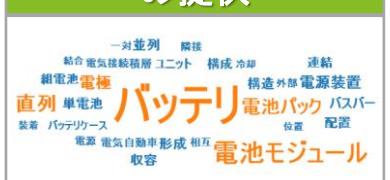
Tu09.交流・直流の変換



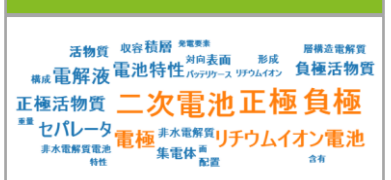
Tu10.エネルギーの変換



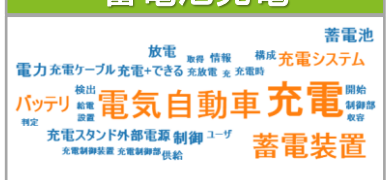
Tu11.電池モジュールの提供



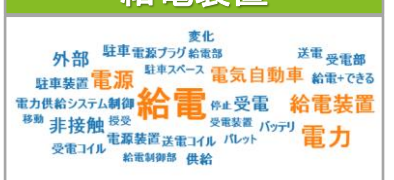
Tu12.二次電池の構成



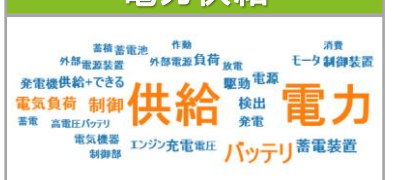
Tu13.電気自動車の蓄電池充電



Tu14.非接触受電など給電装置

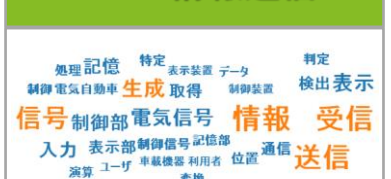


Tu15.外部への電力供給



Tu16.空調などの冷却・加熱

Tu17.情報通信



Tu18.演算・推定

Tu19.機器の異常検出

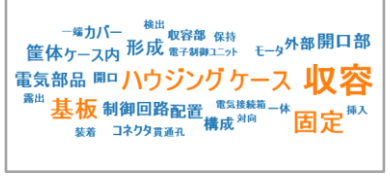
Tu20.操作スイッチ

※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）

特許トピック34個の一覧②

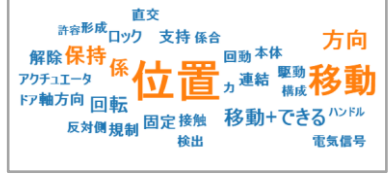
26,419件の特許は、エンジン、動力伝達、モータ、ブレーキ、電力変換、二次電池、充電、情報通信、異常検出、筐体、構成、小型化、安全性などの34個のトピックに集約されました

Tu21.筐体



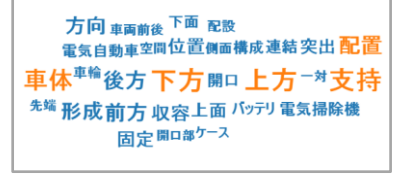
Tu22.表面の形成

Tu23.位置とその移動

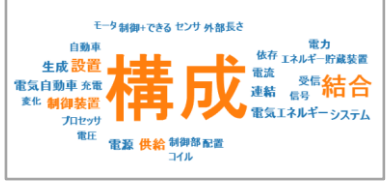


Tu24.配置・位置・方向

Tu25.構成の方位



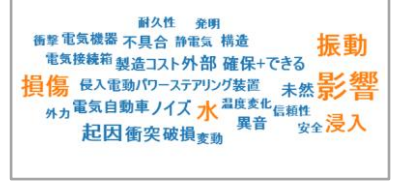
Tu26.構成



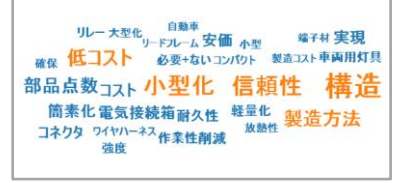
Tu27.接続

Tu28.方法の提供

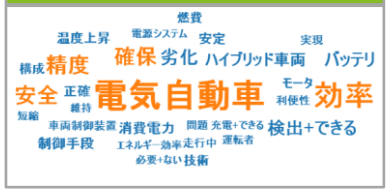
Tu29.損傷や浸水など不具合の防止



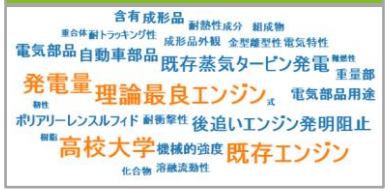
Tu30.小型化・簡素化・低コスト化など付加価値



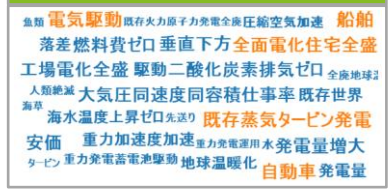
Tu31.効率性・安全性の向上



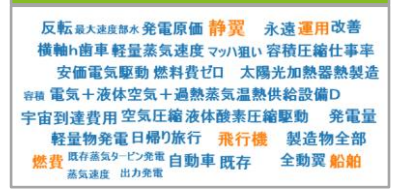
Tu32.既存エンジンへの警鐘・樹脂組成物の提供



Tu33.重力発電の活用による地球温暖化防止



Tu34.タービン発電の出力向上・燃費低減



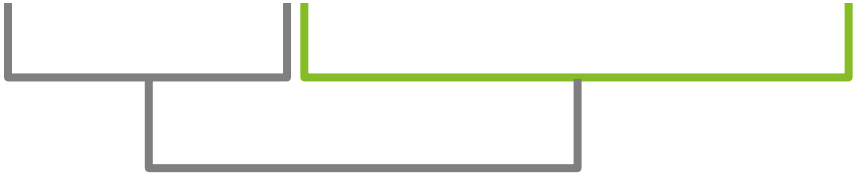
Tu32,Tu33,Tu34は
特定の出願人による
重複した要約内容の
特許から抽出された

※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 （上位5つの単語を赤色で表示している）

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集計・分析を実行することができます

トピックのスコア（該当有無）を紐づけた特許データ

特許ID	出願番号	要約	出願年	出願人	トピック Tu01	トピック Tu02	・・・	トピック Tu34
1	特願2007-XXXX	【課題】電気式変速操作装・・・	2007	A社	1	1		0
2	特願2009-XXXX	【課題】従来の電気自動車・・・	2009	B社	0	1		1
3	特願2012-XXXX	エンジンのための方法及び・・・	2012	C社	0	1		1
4	特願2014-XXXX	【課題】駐車場に設置された・・・	2014	D社	1	0		0
・・・	・・・		・・・	・・・	…	…		…
26,419	特願2016-XXXX	充電ステーションが電気エネ・・・	2016	X社	1	0		1

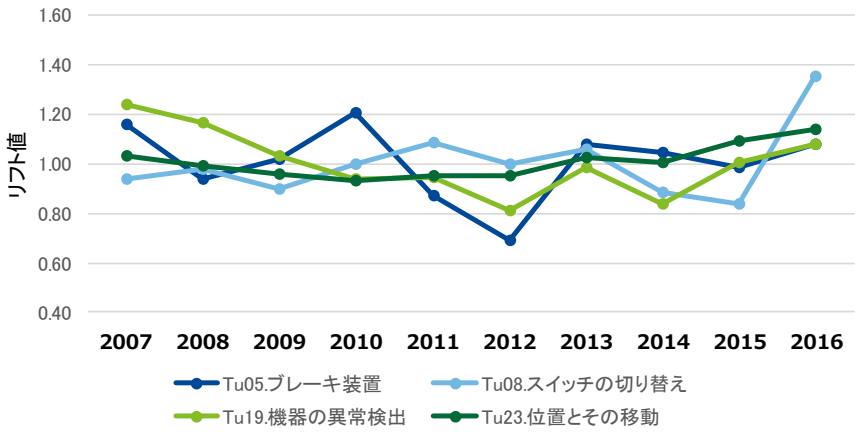


トピック×属性の様々な集計・分析が可能に

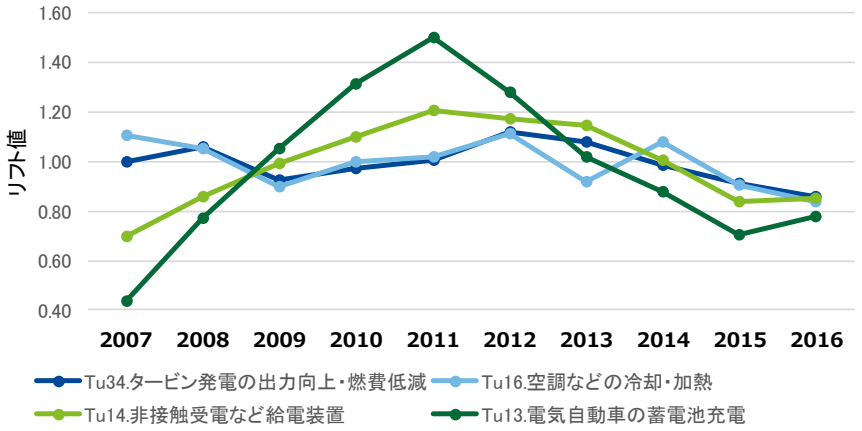
従来の単語ベースではなく集約されたトピックをベースにした分析を実行することで、膨大な特許情報に潜む特徴を分かりやすく理解することができます

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

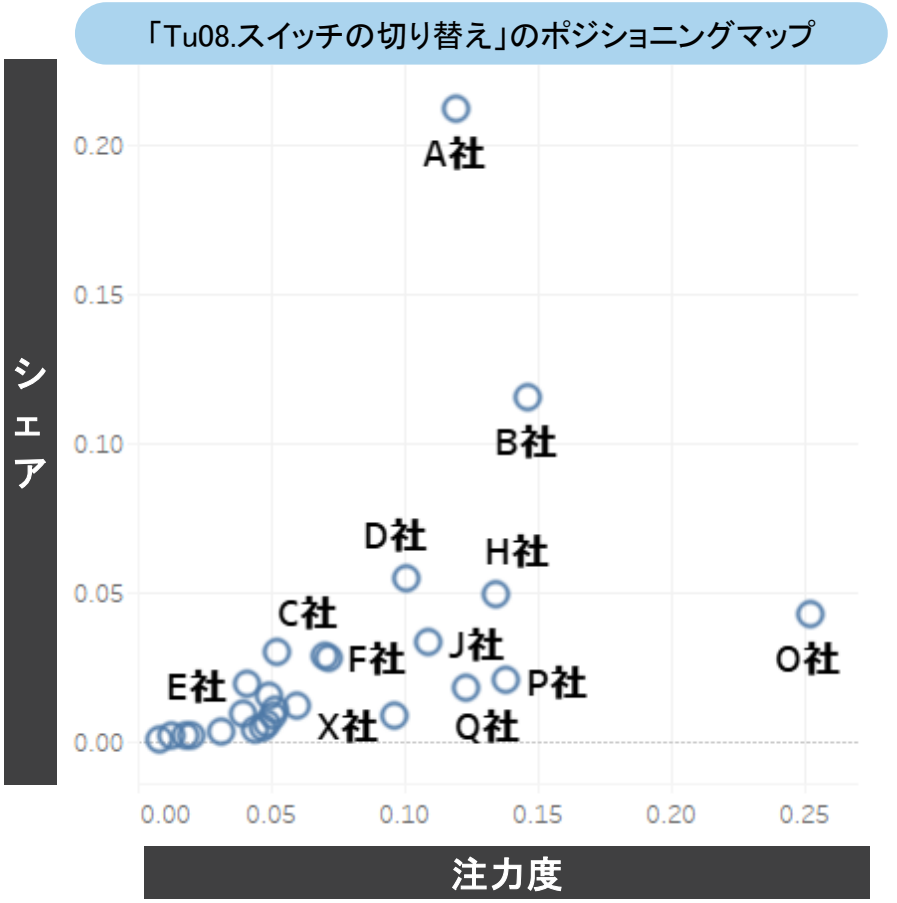


近年下降傾向にあるトピック



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



PLSAを応用した新技術

PCSA（確率的因果意味解析）

【従来手法】 PLSAを用いたトピックの抽出

PLSAでは、テキストマイニングで全体のデータから抽出された単語に基づいた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平均的なトピックを抽出します

テキストマイニング

PLSA

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）

※特許出願中（特願2018-090885）

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

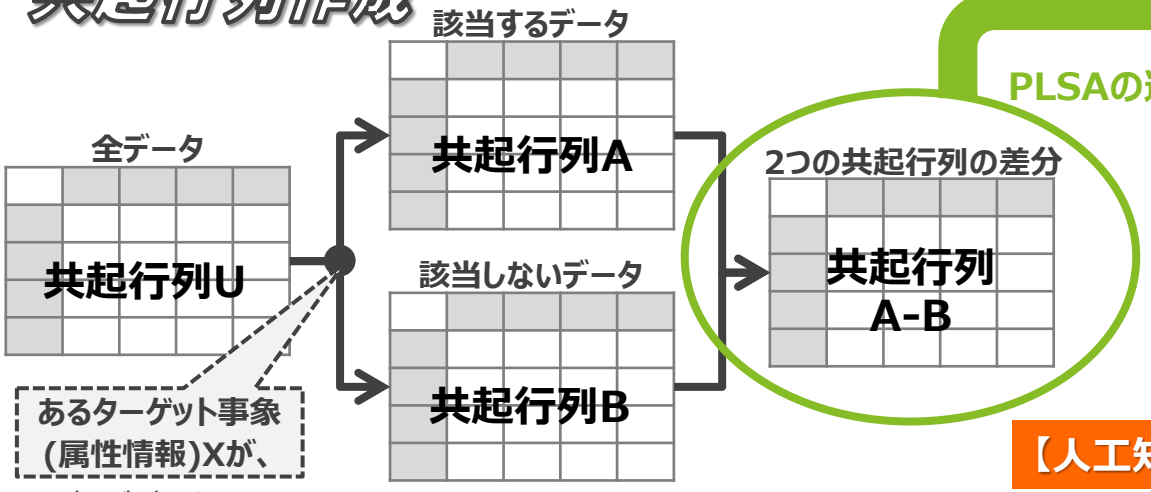
PCSA® (Probabilistic Causal Semantic Analysis：確率的因果意味解析)

テキストマイニング → PLSA

単語抽出

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

共起行列作成



トピック抽出

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

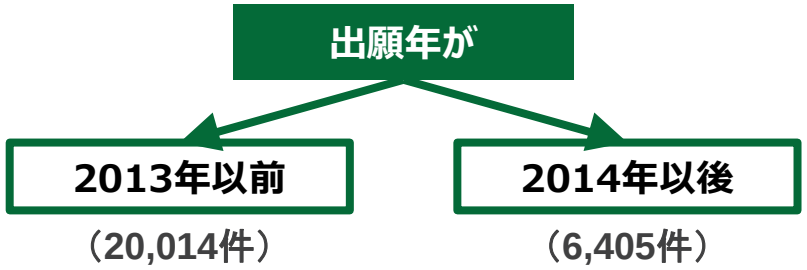
【人工知能学会 2018年度全国大会優秀賞 受賞】

先ほどと同様の電気自動車関連の特許データを対象に、「出願年」をターゲットにPCSAを適用し、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分析対象（先ほどと同様）

- ■ 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
  - ➢ 要約文のテキスト情報

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技術トピックを抽出

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

①2013年以前データ、②2014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

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

単語	係り受け表現				
	電力⇓供給	否⇓判定	モータ⇓駆動	バッテリ⇓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	
…					

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

単語	係り受け表現				
	電力⇓供給	否⇓判定	モータ⇓駆動	バッテリ⇓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	
…					

差分の共起行列の作成

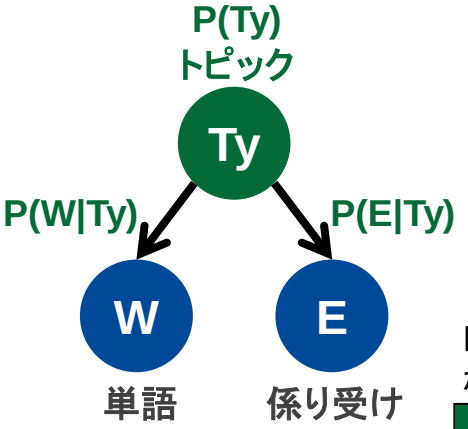
①2013年以前の共起行列と
②2014年以後の共起行列の
差の絶対値を計算した共起行列
を作成する

単語	係り受け表現				
	電力⇓供給	否⇓判定	モータ⇓駆動	バッテリ⇓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	
…					

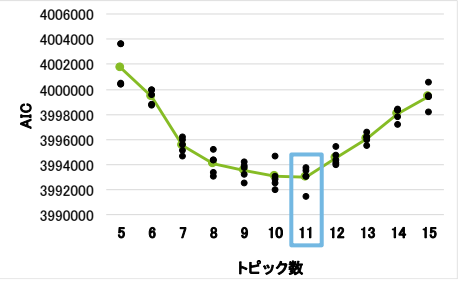
ただし、2つの共起行列は異なる文章数のデータから作成されているので、①2013年以前の共起行列の頻度を2つの文章数の比率（10,723/33,113）で重み調整してから、②2014年以後の共起行列との差を計算する

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



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(Ty)
トピックの存在確率

② P(W|Ty)
トピックにおける単語の所属確率

③ P(E|Ty)
トピックにおける係り受けの所属確率

トピックにおけるP(W|Ty)とP(E|Ty)からトピックの意味を解釈する

トピック Ty05
P(T) = 2.5%

P(W|T)	単語
3.1%	製造
2.2%	二次電池
1.5%	負極
1.5%	正極
1.5%	製造方法
1.2%	エネルギー
1.1%	リチウムイオン電池
1.1%	水素
1.1%	セパレータ
1.0%	電解液
…	…

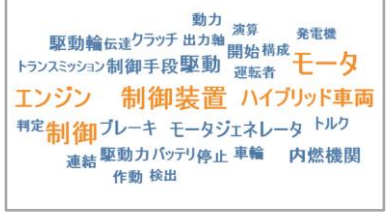
P(E|T)	係り受け
4.5%	水素-製造
3.0%	発電-電気
2.6%	製造方法-提供
2.6%	製造-方法
1.6%	含む-リチウムイオン電池
1.3%	成形品-提供
1.3%	強度-有する
1.3%	表面-形成
1.2%	含む-組成物
1.2%	リレー-スイッチ
…	…

確率の高い構成要素から、Ty05のトピックは「二次電池の製造方法」に関するトピックと解釈できる

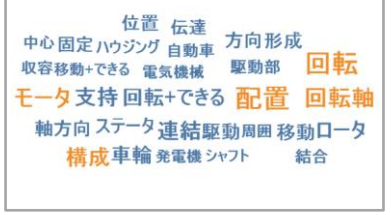
出願年の特徴トピック11個の一覧

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

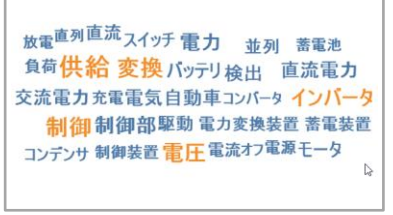
Ty01.エンジン駆動・動力伝達の制御



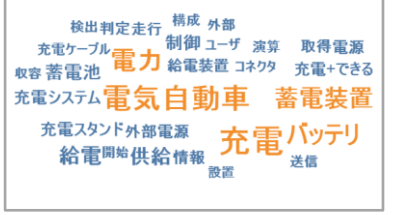
Ty02.モータの回転構成



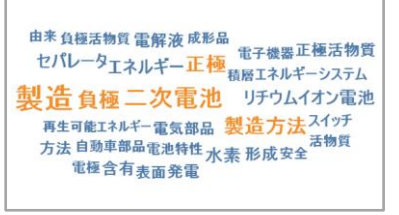
Ty03.交流・直流の変換



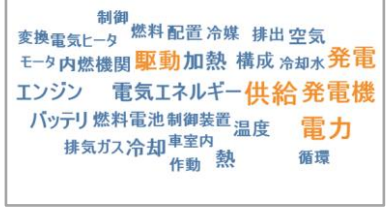
Ty04.電気自動車への充電、給電装置



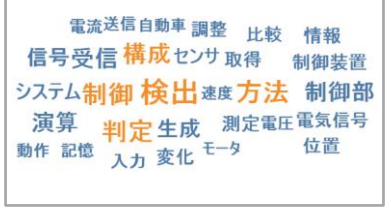
Ty05.二次電池の製造方法



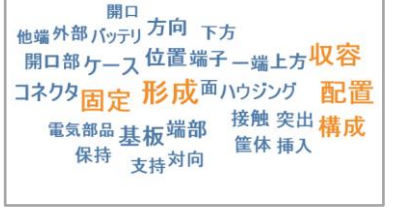
Ty06.空調などの冷却・加熱



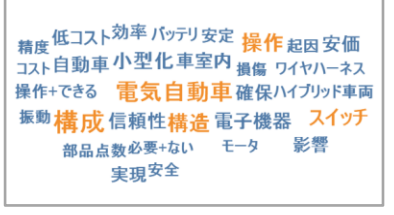
Ty07.情報通信、検出判定システム



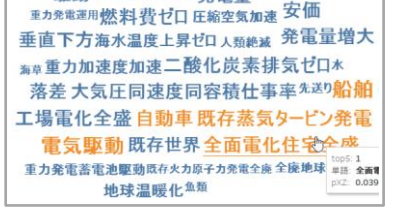
Ty08.形成・配置



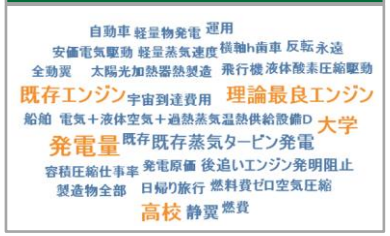
Ty09.小型化・低コスト化・簡素化・操作性向上



Ty10.重力発電の活用による地球温暖化防止



Ty11.既存エンジンへの警鐘、タービン発電・重力発電



Ty10,Ty11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 （上位5つの単語を赤色で表示している）

トピックのスコアデータの作成

全特許データに対して各トピックの該当有無を計算し、各トピックが該当するデータのうち、②2014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（該当有無）を紐づけた特許データ

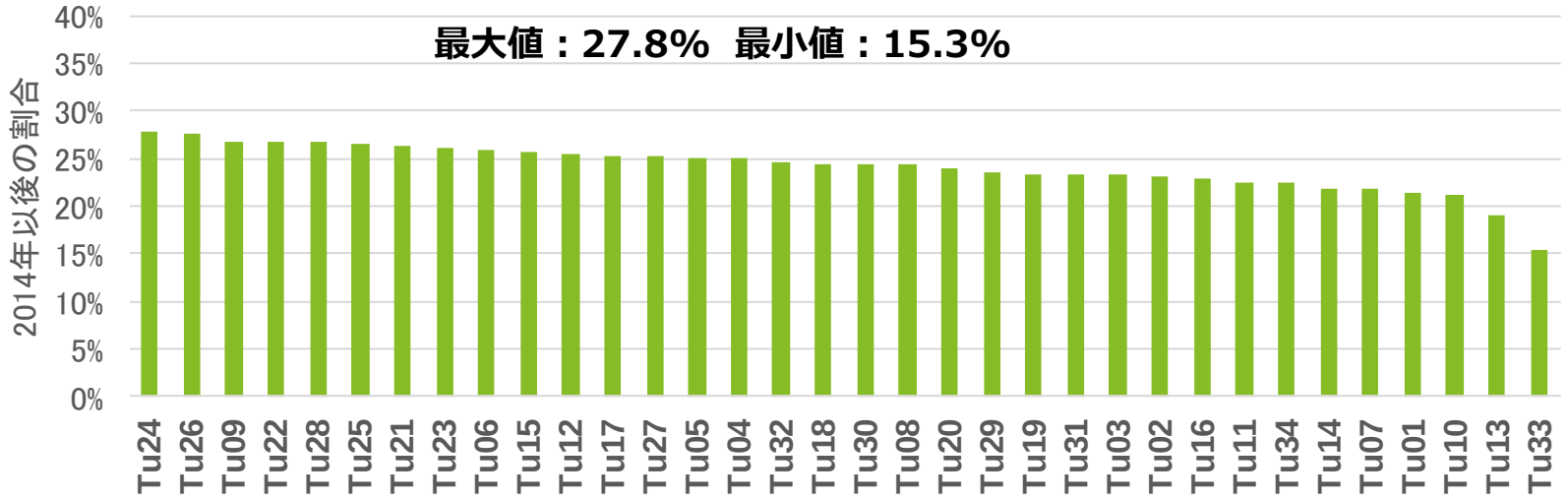
				PLSAによる34個の全体トピック				PCSAによる11個の要因トピック			
特許ID	出願番号	出願年	出願年グループ	トピック Tu01	トピック Tu02	…	トピック Tu34	トピック Ty01	トピック Ty02	…	トピック Ty11
1	特願2007-XXXX	2007	①2013年以前	1	1		0	0	1		1
2	特願2009-XXXX	2009	①2013年以前	0	1		1	0	0		1
3	特願2012-XXXX	2012	①2013年以前	0	1		1	0	1		0
4	特願2014-XXXX	2014	②2014年以後	1	0		0	1	0		0
…	…	…	…	…	…		…	…	…		…
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データのうち、②2014年以後の割合を比較し、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

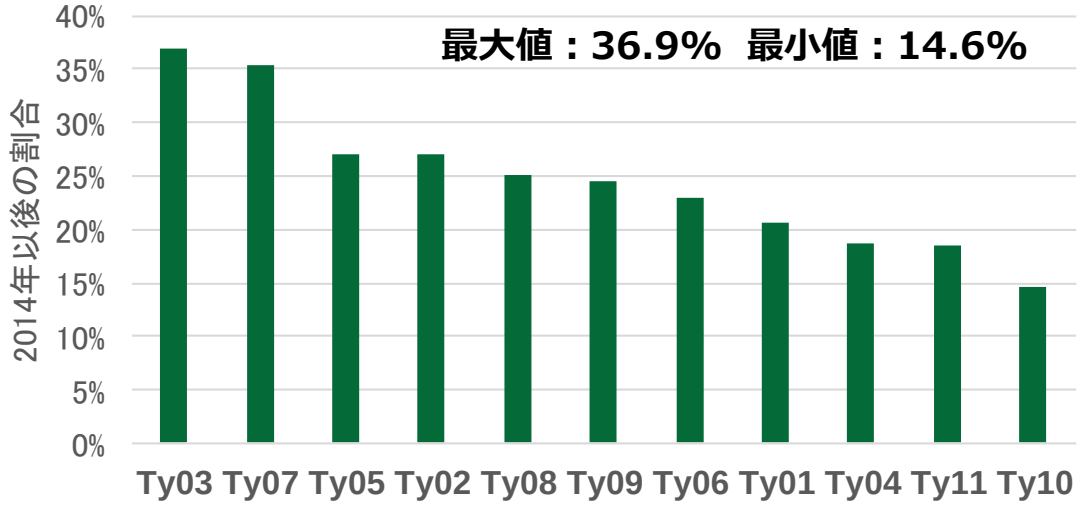
PLSAとPCSAによるトピックの特徴の比較

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



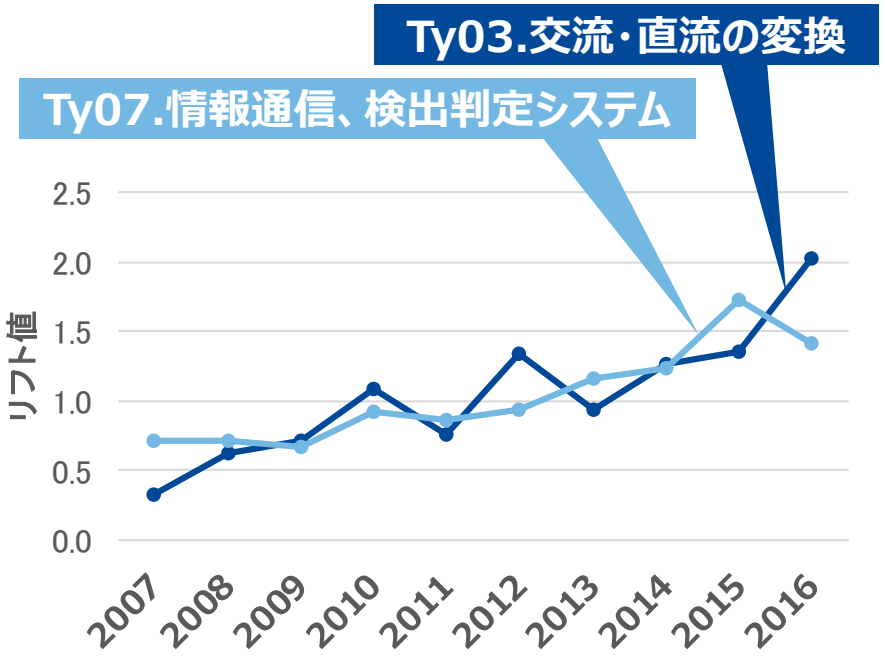
PCSAによる要因トピック



- 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
- 全体のトピックではおおむね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
- 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(Ty03)、データに基づいた運転者のアシスト技術(Ty07)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



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

- 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

Ty03.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。

Ty07.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。

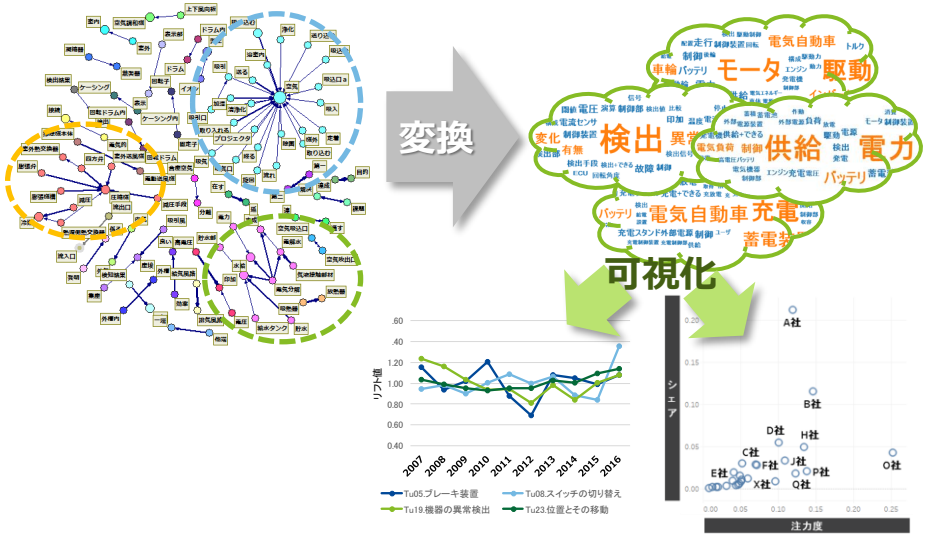
まとめ

トピックモデルというAI技術を応用することで、膨大なテキストデータに潜む特徴をトピックをベースにしたシンプルな可視化で把握し、ビジネスに有用なインサイト獲得が期待できます

トピックモデル　PLSA

テキストデータ全体を表す平均的なトピックを理解し、そうしたトピックの特徴を様々な分析軸で探索する

テキストマイニング → 全体の共起行列 → PLSA

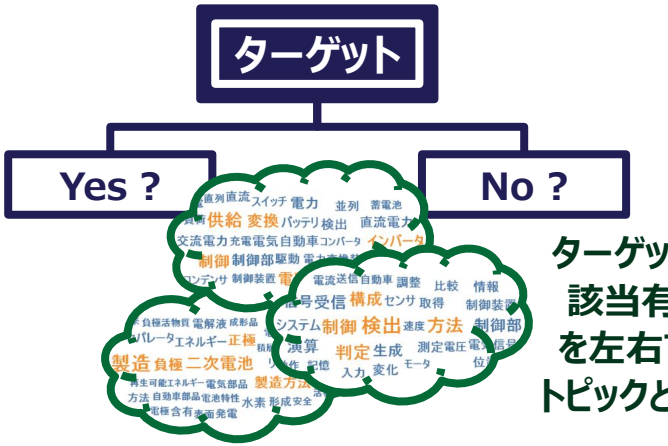


膨大な単語をトピックに変換して特徴を把握できる

PLSAを応用した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？

確認したい対象の要因となるトピックを抽出できる

PCSAの適用対象例

PCSAの技術は様々な業務のテキストデータに適用でき、効果的なビジネスアクションの検討を支援します

口コミ
評価得点に影響を与える
口コミのトピックとは？

アンケート
顧客満足度に影響を与える自由記述トピックとは？

コールセンター履歴
解約・退会に影響を与える
問い合わせトピックとは？

特許文書
近年出願が伸びている
技術トピックとは？

営業日報
契約成立に影響を与える
営業活動トピックとは？

有価証券報告書
業績指標に影響を与える
定性的なトピックとは？

エントリーシート
選考通過に影響を与える
自己PRトピックとは？

診療記録
検査指標に影響を与える
定性的なトピックとは？

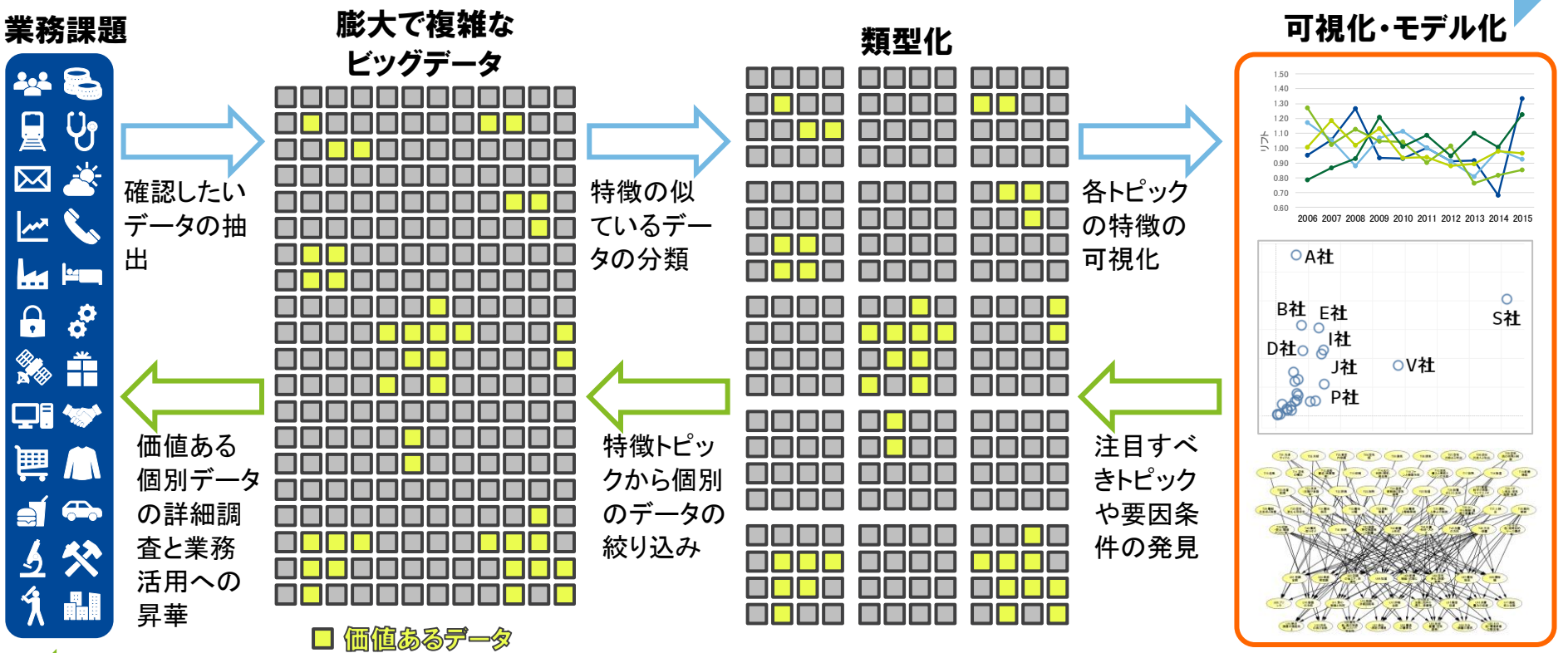
問題発生レポート
問題の発生に影響を与える報告記録トピックとは？

ビッグデータからインサイト獲得のためのダブルアプローチ

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

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

マクロ探索アプローチ



ミクロ探索アプローチ

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

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

Analytics Design Lab