



Analytics Design Lab

2019年度 人工知能学会全国大会

differential PLSA

テキスト情報の典型的なトピックではないより個性的なトピックの抽出

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

2019年6月7日

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

株式会社アナリティクスデザインラボ

企業におけるデータ活用を支援するコンサルティング会社です。



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

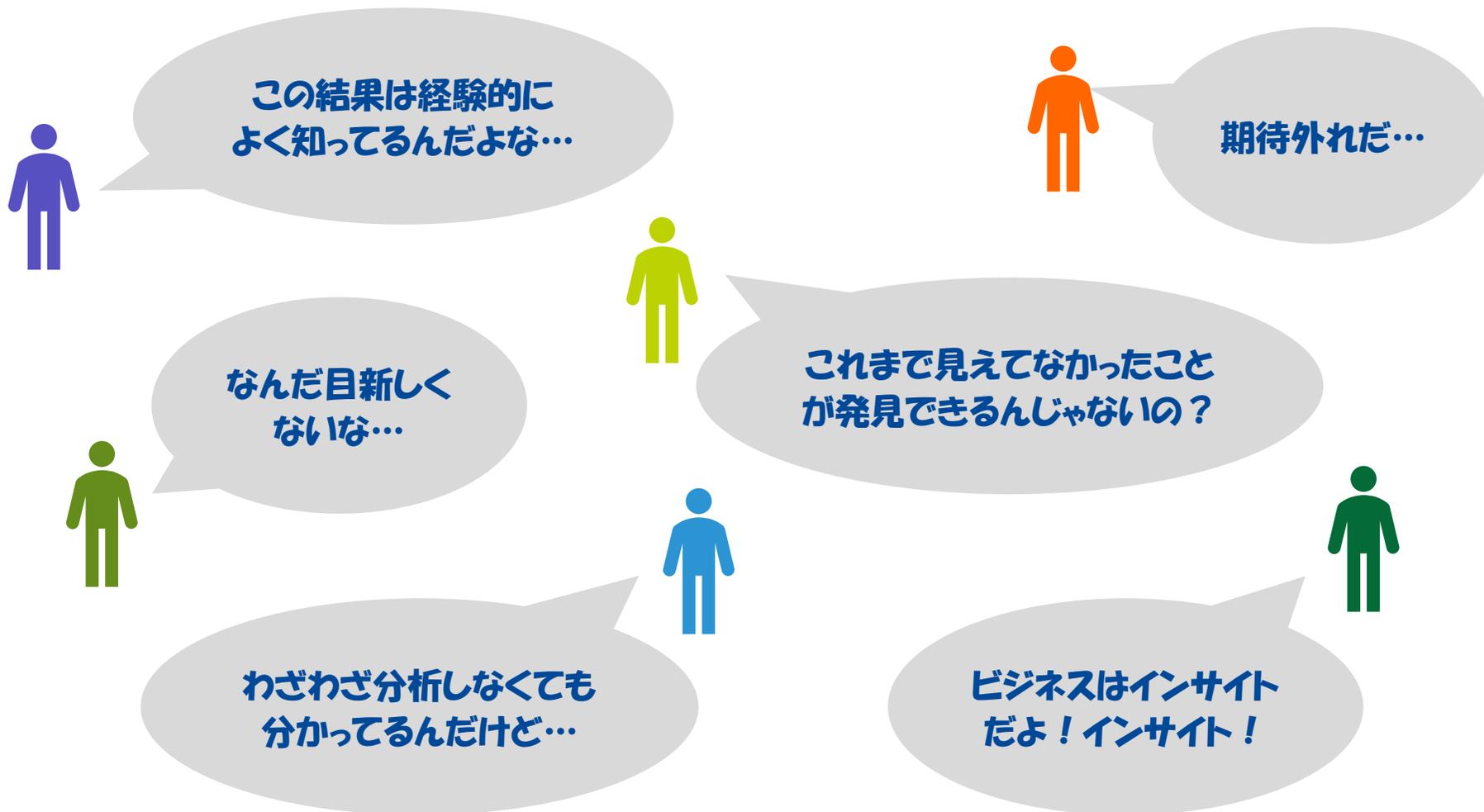
設立	2017年6月1日
事業内容	<ul style="list-style-type: none">● 企業におけるデータ活用のコンサルティング● データ分析技術の研究開発
資本金	5,000,000円
所在地	東京都中野区東中野1-58-8-204

野守 耕爾

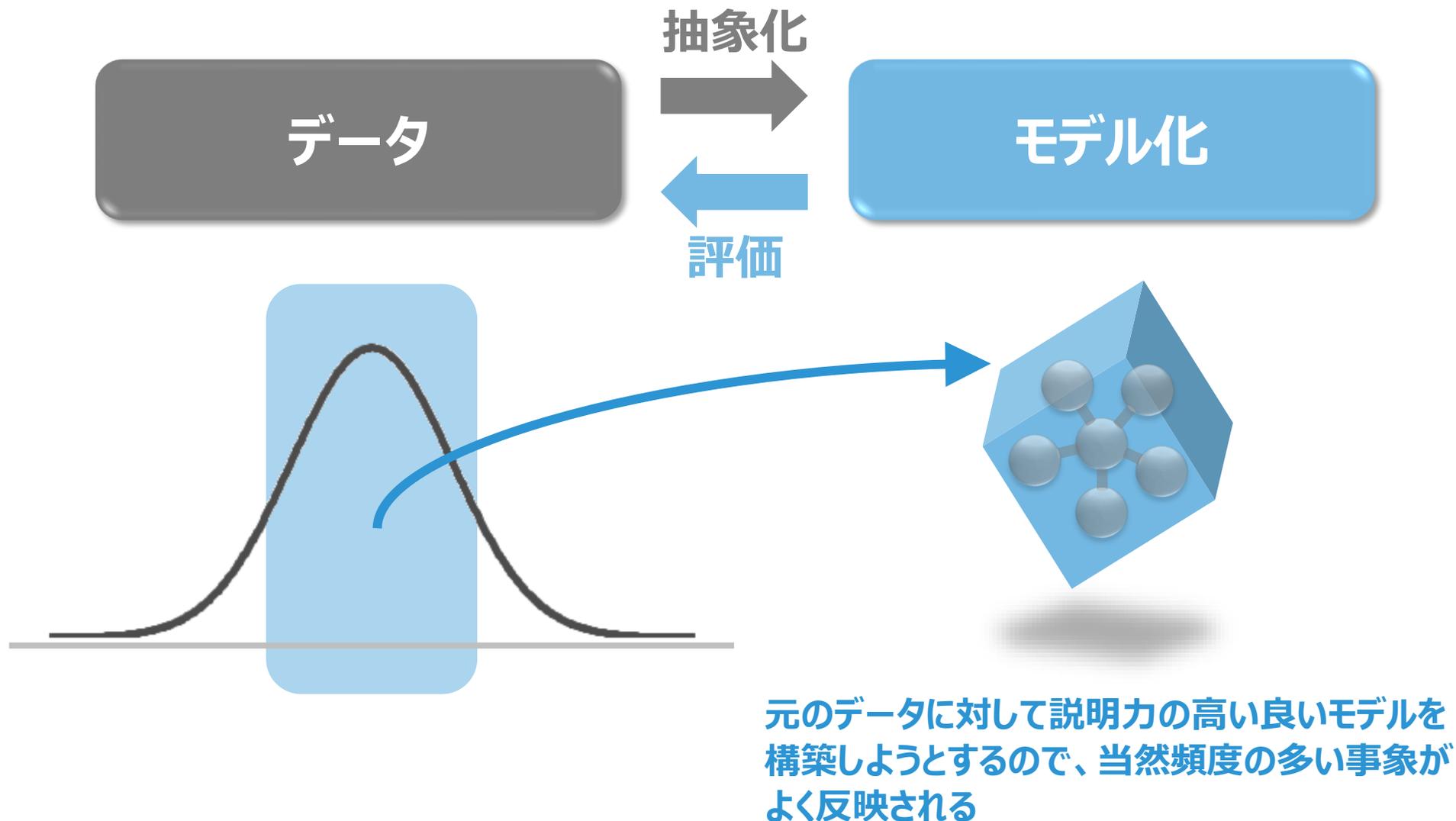


- 2012年3月
早稲田大学大学院 創造理工学研究科
経営システム工学専攻 博士課程修了
博士(工学)
 - 人間行動の計算モデルの開発を研究
- 2012年4月～(技術研修生としては2008年～)
独立行政法人産業技術総合研究所
デジタルヒューマン工学研究センター 入所
 - センシング技術を応用した子どもの行動計測と人工知能技術を応用した行動の確率モデルの開発を研究
- 2012年12月～
デロイトトーマツグループ 有限責任監査法人トーマツ
デロイトアナリティクス 入所
 - データサイエンティストとしてビッグデータを活用したビジネスコンサルティング及び分析技術の研究開発に従事
- 2017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

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



そもそもデータをモデル化すると、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が多い事象≒経験的によく知っていることが優先的に表現されます



テキストデータから典型的なトピックではないより個性的なトピックを抽出するPLSAの新しい手法を提案します

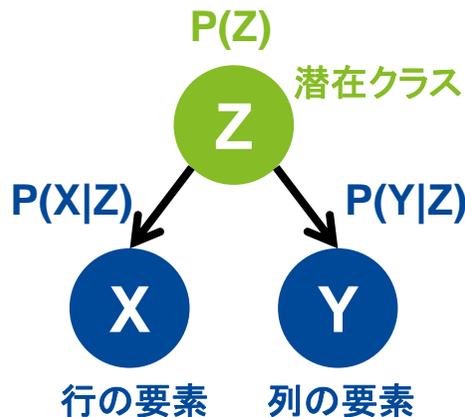


バッテリーのみを駆動源とする電気自動車の、走行中における不要な停車を防止することを可能とした電動車両のバッテリー制御装置を提供する。バッテリー制御装置は、バッテリーとモータとの間を接続する高電圧回路と、高電圧回路の接続と遮断を切り替えるメインコンタクタと、バッテリーを構成する複数の電池セルの電圧を検出する電圧センサと、電圧検出センサが検出したセル電圧が電池セルの過充電を検出するための過電圧検出閾値Xよりも大きい場合に、メインコンタクタにより高電圧回路を遮断するバッテリーマネジメントユニットから構成される。

Mining

電圧センサ、電圧検出センサ、過電圧検出閾値X、メインコンタクタ、バッテリーマネジメントユニット

比較電圧、方向、接続、反対側、電流、保持、入力、停止、制御、検出、形成、電力方法、システム、追加、充電、ハイブリッド車、変換、インバータ、駆動、発電機、電圧、変換、生成、移動、蓄電、接触、構造、変化、温度、開始、安全、情報、ロータ、装置



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

共起行列 $n(X_i, Y_j)$

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

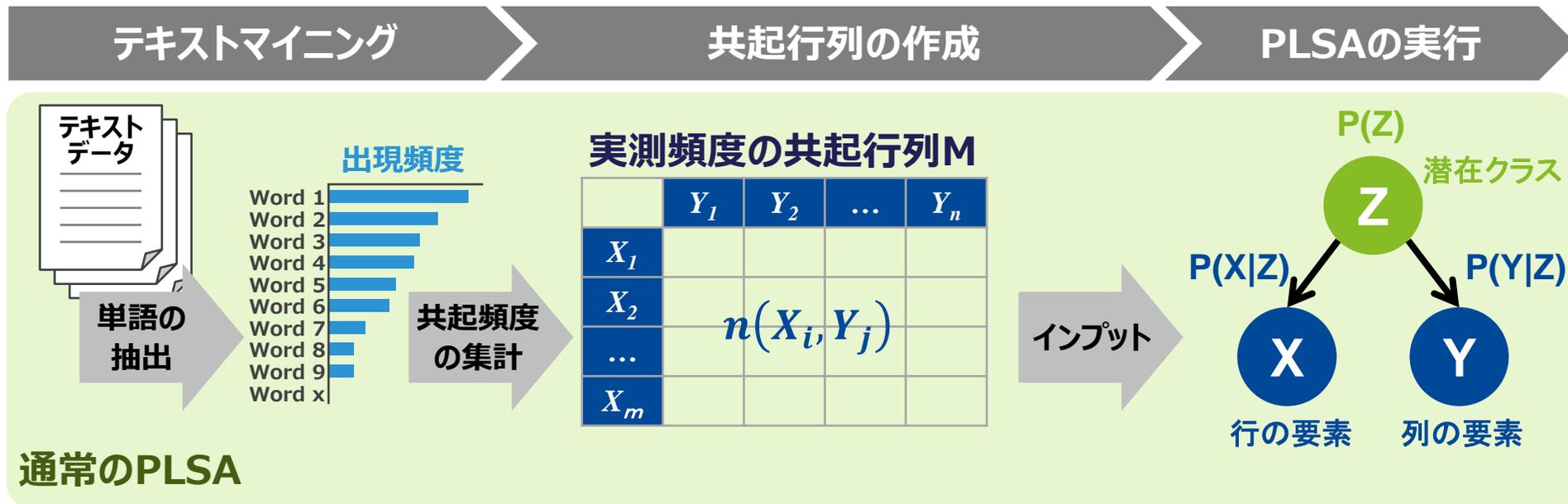
通常は典型的なトピックが抽出されやすいが...

電気自動車、充電、蓄電、バッテリー、電圧、検出、異常、判定、変換、制御装置、二次電池、正極、負極、リチウムイオン、電池モジュール、エネルギー、電気エネルギー、発電機、電力、回転、蓄電、回収、蓄電池、発電、運動工

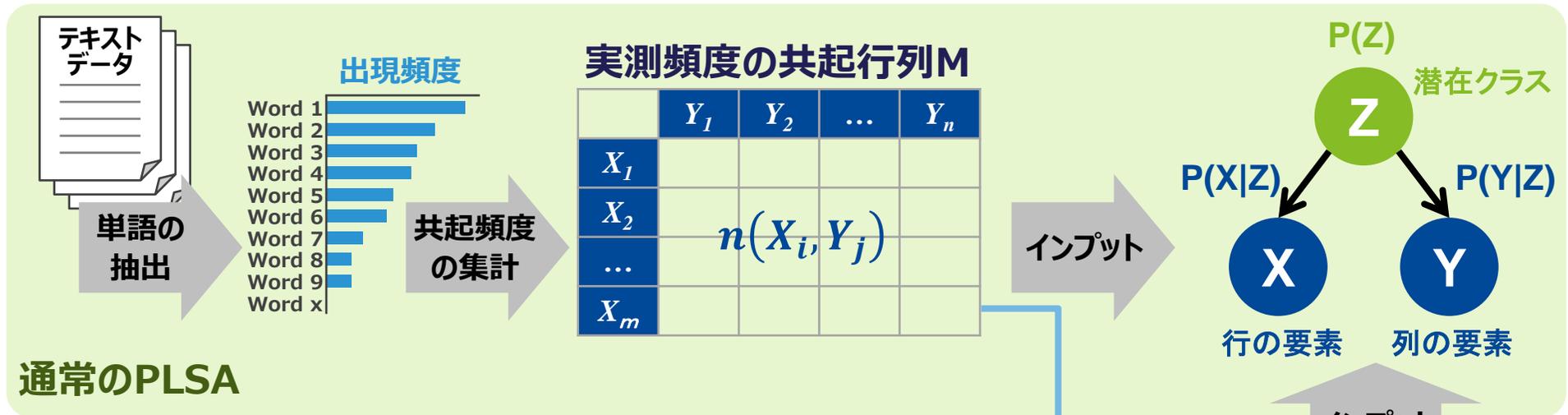
より個性的なトピックの抽出へ

differential PLSA

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



diff-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

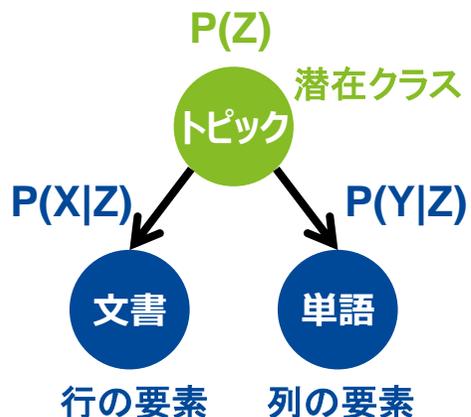
- PLSAを適用した確率解では、**頻度が高い要素に高い確率が割り当てられる**傾向にある
- 結果として抽出されるトピックは**典型的なもの**になってしまう

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

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い共起行列となり、解釈のしやすいトピックを抽出できます

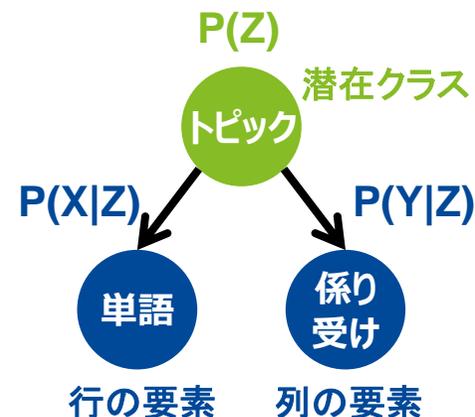
一般的なPLSAの共起行列



	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

本研究の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
- *diff*-PLSAのように共起行列間の比率を取る効果も大きい
- 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

differential 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、通常のPLSAとdiff-PLSAでそれぞれトピックを抽出し、結果を比較します

データの抽出条件と分析対象

- 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
 - 要約文のテキスト情報



要約文の例

発明の名称
移動体用非接触充電装置および移動体用非接触充電方法
要約
【課題】受信部の設置に制約のある移動体に効率的に非接触で電力を伝送すること。【解決手段】移動体用非接触充電装置は、電気自動車に設けられる受電装置と、受電装置と非接触で設けられる送電装置とを備えている。受電装置は、電気自動車の所定の箇所に分散して配置され、外部からの電力を受信するための複数の受電コイルを備える。送電装置は、複数の受信コイルに対応して電気自動車の外部の所定位置に配置され、電力を送信するための複数の送電コイルを備える。複数の送電コイルから複数の受電コイルに非接触で電力を伝送し、二次電池を充電する。受電コイルを複数設けるため、設置面積または設置形状に制約がある場合でも、効率よく電力を伝送できる。

※国内の特許の要約は400字以内の字数制限がある

通常のPLSAによるトピック抽出

テキストマイニングで単語と係り受け表現を抽出し、単語×係り受けで構成される実測の共起行列にPLSAを適用することで、約2万6千件の特許の全体を表現するトピックを抽出します

テキストマイニングの実行

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
...	...

※ここでの頻度は出現文章数ではなく出現データ件数

共起行列の作成

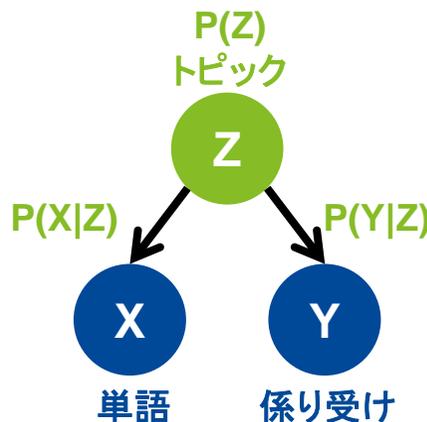
「単語×係り受け」の共起行列(文章単位で同時に出現する頻度のクロス集計表)を作成する

	係り受け表現				...
	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	
...					

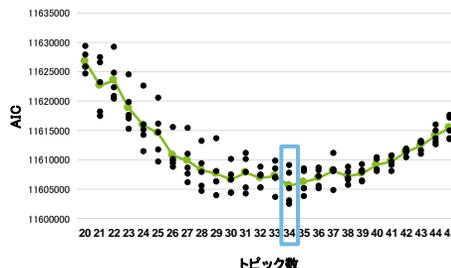
単語(名詞): 3,020語
 係り受け: 2,128表現
 ※データ頻度20件以上を対象

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 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、Z13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

通常のPLSAで抽出された特許トピック34個の一覧②

通常のPLSAでは、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個のトピックが抽出されました



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（上位5つの単語を赤色で表示している）

differential PLSAによるトピック抽出

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diff-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
			電力↓供給	否↓判定	モータ↓駆動	バッテリー↓充電
単語	総頻度 n(X)	5,880	118	33	36	33
	構成	5,092	268	73	108	85
	制御	4,604	69	2	29	8
	配置	5,188	239	61	494	58
	モータ

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

※総文章数: 229,598件

diff-共起行列の作成

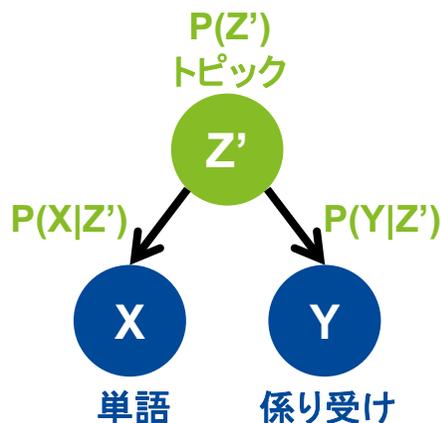
2つの共起行列の各共起ペアにおいて、期待共起頻度に対する実測共起頻度の比率の対数を取ったdiff-共起行列を作成する

		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	...
単語	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	

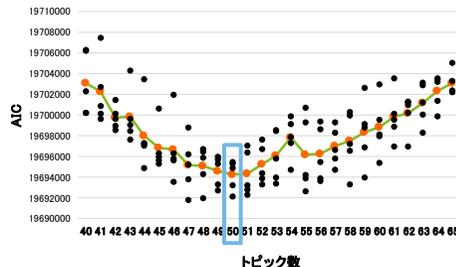
※値が負数となるものは"0"に置換する

PLSAの実行

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、Z'20のトピックは「充電の接続」に関するトピックと解釈できる

diff-PLSAで抽出された特許トピック50個の一覧①

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'01.エンジン制御</h3> <p>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</p>	<h3>Z'02.動力伝達</h3> <p>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 動力逆送車装置 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</p>	<h3>Z'03.差動機構などを備えた動力伝達の制御</h3> <p>変速制御運転状態 差動部 変速比 駆動部 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</p>	<h3>Z'04.回転運動</h3> <p>駆動軸 同軸 回転カ一方 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</p>	<h3>Z'05.ロータ・ステータなどモータの構成</h3> <p>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</p>
<h3>Z'06.モータ制御(トルク制御や回転数制御など)</h3> <p>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 電流センサ演算 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</p>	<h3>Z'07.油圧ポンプなどを利用したモータ駆動</h3> <p>電力 供給 車輪 走行用 後輪 駆動 エンジン 動力トルク 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</p>	<h3>Z'08.ブレーキ</h3> <p>押圧力操作部材 ピストン 制御操作部材 摩擦部材 ツパバ 運転者付与 車輪 モータ シリンダ 装置 制動力 ブレーキ操作 液圧 スレーブシリンダ ブレーキ液圧 入力装置 マスタシリンダ ブレーキ 電気ブレーキ ブレーキ液圧 制動力 トルク ブレーキペダル 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</p>	<h3>Z'09.状態に応じた制御、運転者の操作補助</h3> <p>点滅 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 シフトレバー処理 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 解除 キースイッチ 開状態 自動的 検出 警告 切替故障 開始</p>	<h3>Z'10.コンバータとバッテリー昇降圧</h3> <p>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 再生電力 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 オルトスターター 電源装置 電圧コンバータ バッテリー 発電機 電力 直流</p>
<h3>Z'11.直流と交流の電力変換</h3> <p>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 直流電圧 整流 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 昇圧 交流 供給 変換電力変換</p>	<h3>Z'12.回転力などの電気エネルギー変換</h3> <p>海水 蓄電 水車 太陽光 上下動 流体 風 運動エネルギー 風車 カ回転 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</p>	<h3>Z'13.エネルギー効率の向上</h3> <p>快追損失 技術 走行+できる 精度 回避 実現 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</p>	<h3>Z'14.発電と蓄電</h3> <p>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</p>	<h3>Z'15.電池モジュールの提供</h3> <p>内蔵 構造 電極 相互 積層方向 バッテリー 隣接 スタック 単電池 ベースプレート 横方向 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 パスバー 下端 片側 構築 直列外形</p>
<h3>Z'16.燃料電池</h3> <p>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</p>	<h3>Z'17.二次電池の構成</h3> <p>外装部材 含リチウムイオン バッテリー 介在 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 活物質 リチウムイオン電池 両面 正極 負極 電解要素 電極素子 負極活物質 電気抵抗 負極集電体 積層 外装体</p>	<h3>Z'18.バッテリーの充放電</h3> <p>充 電源 警報建物 放電 給電 給電装置 蓄電池 負荷電力変換装置 授受 充放電装置 系統電源 停電時 指示 電力供給システム プレーガ 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 受受信 電力 開閉器</p>	<h3>Z'19.充電システム</h3> <p>充電 充電電流 駐車場 充電動作 駐車 利用者外部電源 記位 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</p>	<h3>Z'20.充電の接続</h3> <p>充電電線 商用電源 充電時 電力変換部 車両側コネクタ 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</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

<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 バレット 非接触給電システム 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 コイル 給電受電装置 給電部 非接触給電装置</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 警告音 回転角度 異常 制御回路有無 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witching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 パラメータ 記憶手段 制御手段 発電電力 関値 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 バッテリー残量 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 ストッパ ストッパ 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 没入 操作時 変形 安価 勝手 力 駆動操作</p>
<h3>Z'31.車両用灯具</h3> <p>貫通 近接 LED おじ クランプ 反射 出射 発熱体 車両用灯具 灯室内 ハウジング 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 放熱部材/バルブ 入射 固定部材 透過 光 基板</p>	<h3>Z'32.掃除機</h3> <p>車輪連通 前方 開口 下面 前面 転動 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 上方 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 較大型車両 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 車輻方向 電池モジュール 搭載構造 車体 フレーム クロスメンバ 前輪 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 アンコイル位置 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 係 押圧解除 ロックレバー 車両前方 規制 許容 回転</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で抽出された特許トピック50個の一覧③

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 前車列 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
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)

材料誘電率電圧金型剛性
耐薬品性電気部品用途 成形品 外観 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 耐トランッキング性 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 外観 成形性 脆断性
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 異常環境下 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 遮断動電 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 電池バック バイパス 電気接続
一側面 トレーラ バッテリ 電荷 並列 正

Z'49.タービン発電と船舶・飛行機への応用

既存 日帰り旅行 運用 出力発電 燃料費ゼロ
安価電気駆動 製造物全部 発電量 発電原価宇宙到達費用
液体燃料圧縮駆動 振動伝達 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 高速推進水 自動車 反転

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 全端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
自然現象高効率圧縮空気加速 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 (上位5つの単語を赤色で表示している)

通常のPLSAと*diff*-PLSAの結果の比較

【比較1】diff-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

P(XZ)	n(X)	X: 単語	P(YZ)	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	1.5%	59	電気信号-基づく
...

P(XZ)	n(X)	X: 単語	P(YZ)	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	1.4%	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

P(XZ)	n(X)	X: 単語	P(YZ)	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	2.0%	92	備える-ブレーキ
1.3%	111	操作量	2.0%	92	車両-ブレーキ
...

P(XZ)	n(X)	X: 単語	P(YZ)	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	非接触-受電
1.8%	124	受電部	1.7%	34	供給-給電装置
...

※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】diff-PLSAのみで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

P(XZ)	n(X)	X: 単語	P(YZ)	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

P(XZ)	n(X)	X: 単語	P(YZ)	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

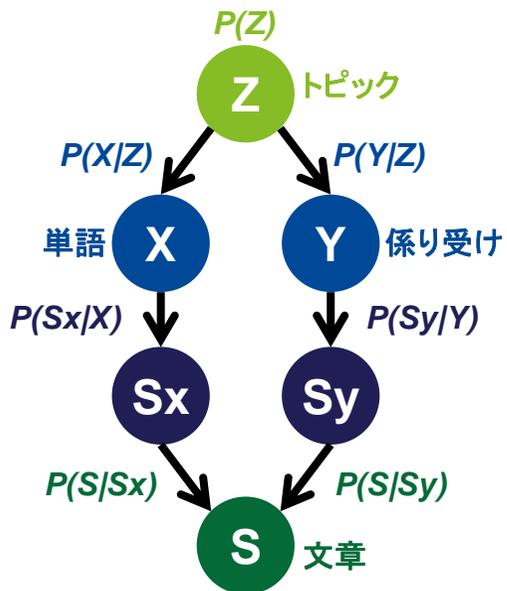
充電スタンドへの電気自動車の有効なナビゲーションを行う。本発明は、複数の充電スタンドの位置情報を記憶するデータベースと、電気自動車と通信して、電気自動車が搭載する電池の種別、エネルギー残量、エネルギー効率、現在位置を受信し、現在位置から電池残量およびエネルギー効率によって到達可能な充電スタンドを見つけ、その位置情報と、その充電スタンドが保持するエネルギー残量の情報を電気自動車に送信する充電スタンドロケータとを備える。

全特許データに対して各トピックのスコア(該当有無)を計算することもできます

文章単位
のスコア

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

- 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 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	...	T50
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	...	T50
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10に設定する

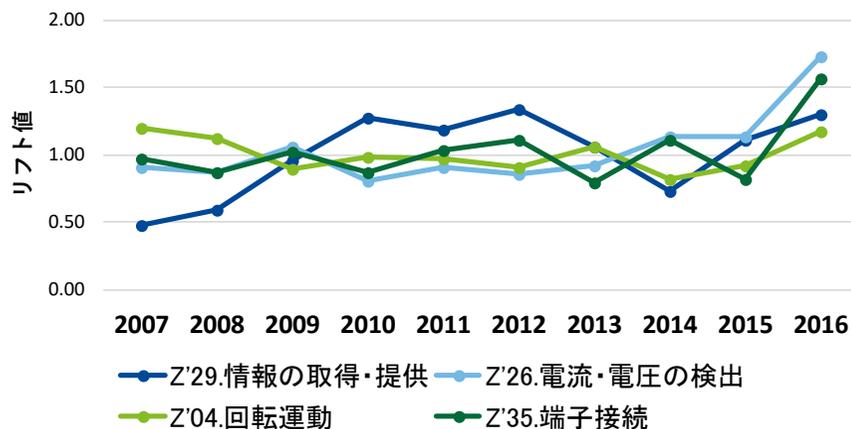
特許ID	T01	T02	T03	...	T50
1	1	0	1		0
2	0	1	0		1
...					

トピックをベースとした特徴分析

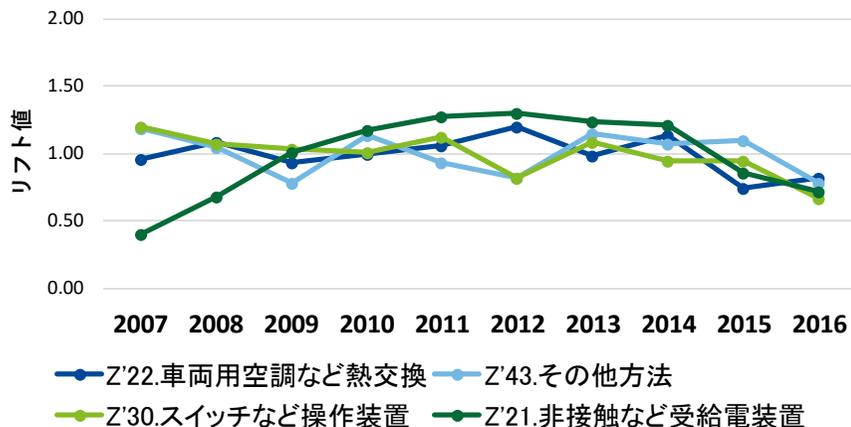
トピックのスコアを出願年や出願人の属性軸で分析することで、技術トレンドや競合他社のポジショニングを可視化でき、技術戦略を検討する上で効率的にインサイト獲得が期待できます

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



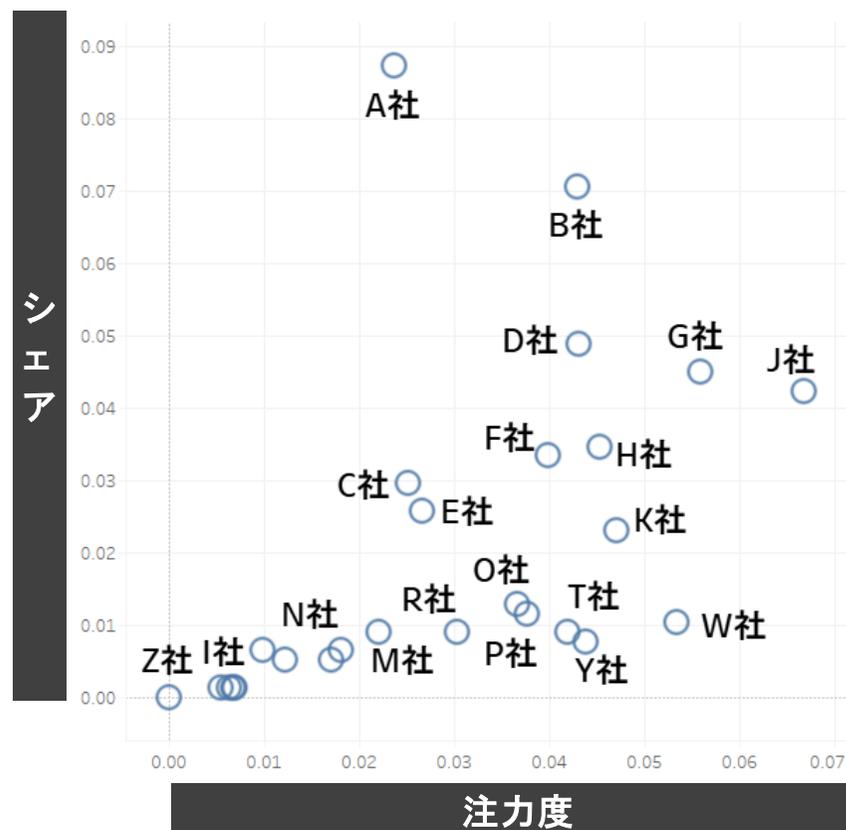
近年下降傾向にあるトピック



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'22.情報の取得・提供」のポジショニングマップ



まとめ

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



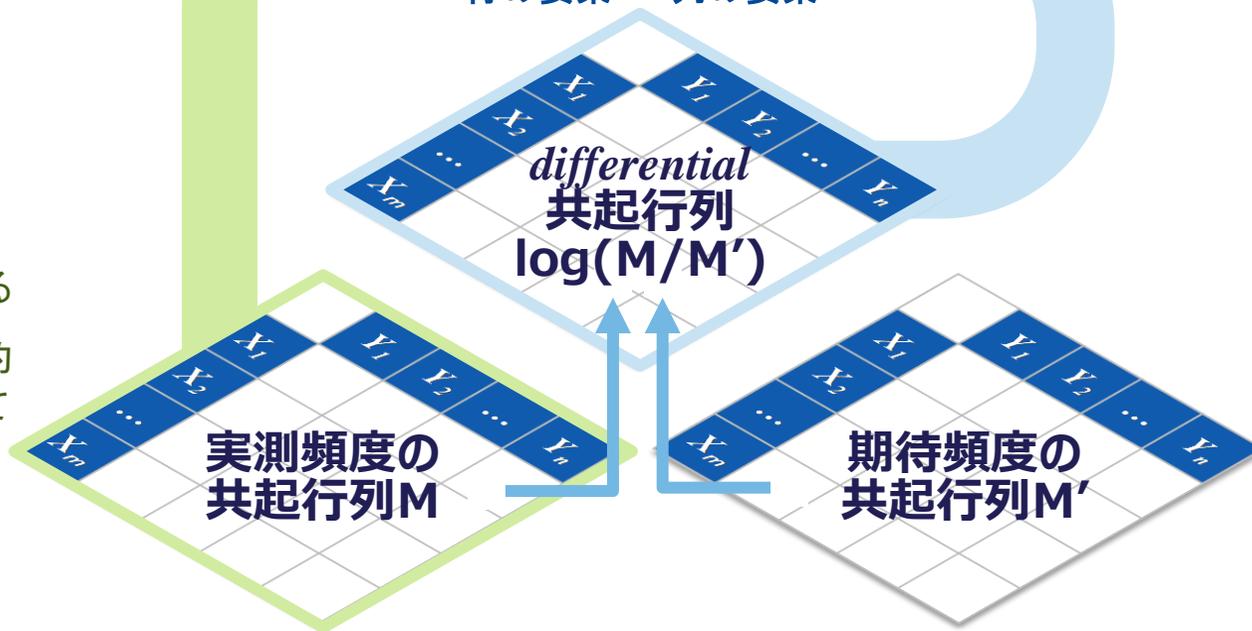
行の要素 列の要素

インサイトを得る



differential PLSA

- 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



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

