



Analytics Design Lab

ビッグデータ活用展2018 spring  
株式会社NTTデータ数理システム ブース内セミナー

## 特許文書データにAI技術を応用した技術戦略に資する特徴分析

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5月10日、11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"><li>● 企業におけるデータ活用のコンサルティング</li><li>● データ分析技術の研究開発</li></ul>
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



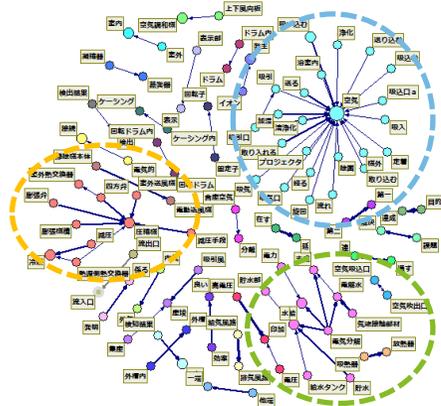
- 2012年3月  
早稲田大学大学院 創造理工学研究科  
経営システム工学専攻 博士課程修了  
博士(工学)
  - 人間行動の計算モデルの開発を研究
- 2012年4月～(技術研修生としては2008年～)  
独立行政法人産業技術総合研究所  
デジタルヒューマン工学研究センター 入所
  - センシング技術を応用した子どもの行動計測と人工知能技術を応用した行動の確率モデルの開発を研究
- 2012年12月～  
デロイトトーマツグループ 有限責任監査法人トーマツ  
デロイトアナリティクス 入所
  - データサイエンティストとしてビッグデータを活用したビジネスコンサルティング及び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

# 人工知能技術を応用した新たな特許文書分析のアプローチ

# これまでの特許文書分析

単語をベースに、あるいは手動でグルーピングしたカテゴリをベースに、全体の出現状況、経年変化、出願人の特徴、課題と解決手段の対応関係などを把握する分析がよく行われます

## 共起ネットワークによる全体像把握

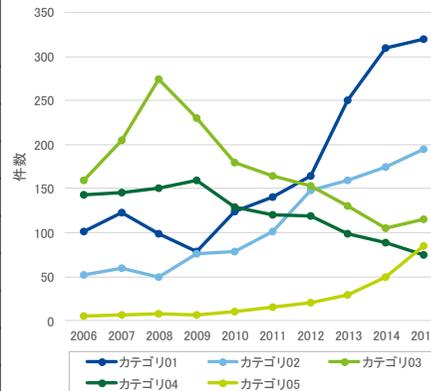


- 単語の共起関係をネットワークで可視化する
- ネットワークのかたまりを見ながら、全体でどのような話題が形成されているのか考察する

## 手動設定したカテゴリのトレンド把握

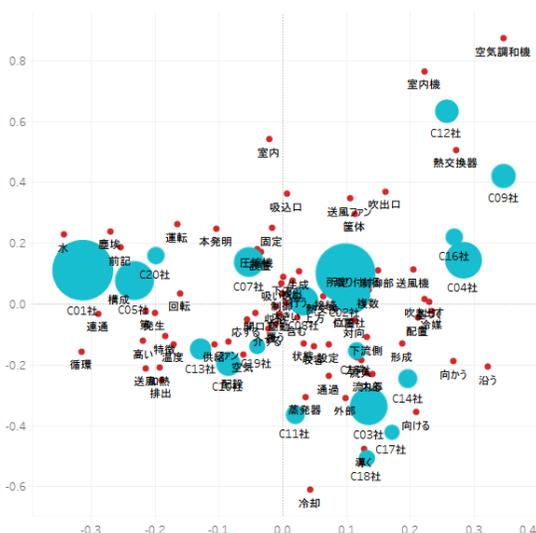
例) 掃除機カテゴリのリスト

掃除機
集塵
集塵容器
吸引力
サイクロン
塵埃->分離
塵埃->吸い込む
塵埃->収容
塵埃->遠心分離



- 抽出した単語を手動でいくつかのカテゴリにグルーピングする
- 各カテゴリの出願年ごとの出現頻度をグラフ化し、トレンドを把握する

## コレスポネンス分析による出願人の特徴把握



- 単語の出現データから共通して現れる特徴的な軸を2つ抽出する
- その2軸による平面上に単語と出願人を同時にマッピングする
- 出願人の周辺に配置された単語群から各出願人の特徴を考察する

## 課題と解決手段のクロス集計による関係把握

課題	解決手段									
	カテゴリ01	カテゴリ02	カテゴリ03	カテゴリ04	カテゴリ05	カテゴリ06	カテゴリ07	カテゴリ08	カテゴリ09	カテゴリ10
カテゴリ01	206	80	71	184	26	47	11	9	43	1
カテゴリ02	208	76	87	182	23	48	9	15	40	2
カテゴリ03	172	74	53	57	31	35	10	21	20	3
カテゴリ04	176	54	37	59	26	46	29	26	9	5
カテゴリ05	85	39	13	23	14	16	5	0	7	2
カテゴリ06	87	53	31	33	59	37	15	24	28	19
カテゴリ07	79	68	82	28	24	12	6	16	18	15
カテゴリ08	32	29	19	1	20	5	17	2	4	2

- 「要約」の【課題】と【解決手段】それぞれに対して出現単語のカテゴリを設定する
- 課題と解決手段のカテゴリのクロス集計をして、用途と技術の関連性を考察する

複数の人工知能技術を組み合わせることで、特許データを単語ベースではなく、客観的に抽出されるトピックベースで解釈し、そのトピックの統計的な関連性を分析できます

単語ベースの分析では  
複雑で考察しにくい

カテゴリの設定が主観的で  
作業負荷も大きい

課題と解決手段の統計的な  
関係を分析していない

単語を賢くクラスタリングする  
人工知能技術

要因関係をモデリングする  
人工知能技術

**PLSA**  
確率的潜在意味解析

文脈を考慮した潜在的なトピック  
(単語の集合)を抽出する

**ベイジアンネットワーク**

多様な要因間の確率統計的な  
因果関係をモデル化する

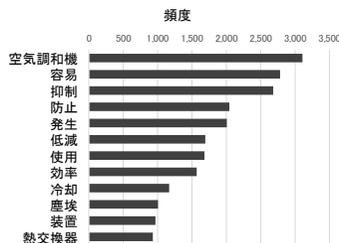
膨大なテキストデータをトピックに変換して解釈を容易にし、テキスト情報内に潜む要因関係をモデル化して、ビジネスアクションに有用な特徴を把握可能にします

# Nomolytics<sup>®</sup>: Narrative Orchestration Modeling Analytics

## テキストマイニング

文章に含まれる単語を抽出し、その出現頻度を集計する

単語抽出



## PLSA

### 確率的潜在意味解析

単語が出現する文脈を学習し、膨大な単語を複数のトピックにまとめる

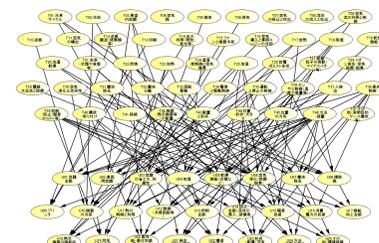
トピック抽出



## 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

モデリング



膨大なテキストデータを人間が理解しやすい形に整理できる

テキスト情報内に潜む複雑な要因関係を構造化できる

条件を変化させたときの結果の挙動をシミュレーションできる

ある事象の発生確率をコントロールする条件を発見できる

# Nomolyticsを適用した特許分析事例

## 「風」「空気」に関する10年分の特許データ30,039件を分析します

### 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～2015年
- 抽出方法
  - PatentSQUAREを使用
- 抽出結果
  - 30,039件



### 分析データの加工

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

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

## 特許要約の【課題】と【解決手段】から用途と技術のトピックを抽出し、トピックのトレンド分析や出願人の特徴分析、また用途と技術の関係分析による新規用途探索を行います

### A. 用途と技術のトピック抽出

データの抽出



- 特許文書の要約文の「課題」と「解決手段」のテキストデータを抽出する
- 「課題」からは用途トピックを、「解決手段」からは技術トピックを抽出する

### テキストマイニング

テキストマイニングを実行して単語と係り受け表現を抽出する

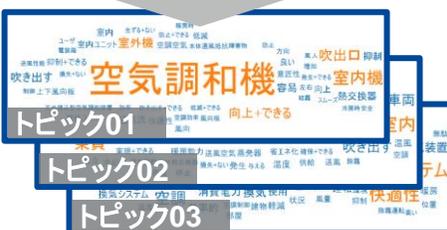
単語	品詞	頻度
空気調和機	名詞	3,106
空気	名詞	2,846
容易	名詞	2,790
抑制	名詞	2,687
...	...	...

係り受け表現	頻度
空気調和機-提供	1,575
効率-良い	1,325
掃除機-提供	545
容易-構成	539
...	...

### PLSA

「単語×係り受け」の共起行列を作成し、これにPLSAを適用してトピックを抽出する

単語	係り受け			
	機提供	空気調和	効率-良い	掃除機-提供
空気調和機	1,578	100	1	
空気	85	144	45	
容易	190	105	67	



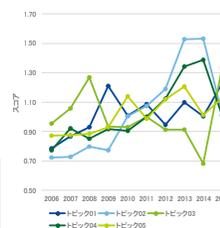
### B. トピックの特徴集計

全特許データに各トピックのスコア(該当度)を計算する

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	...	...	...	...	...	...	...	...

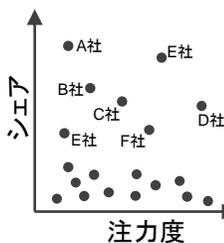
### 出願年集計

トピックスコアを出願年で集計してトピックのトレンドを把握する



### 出願人集計

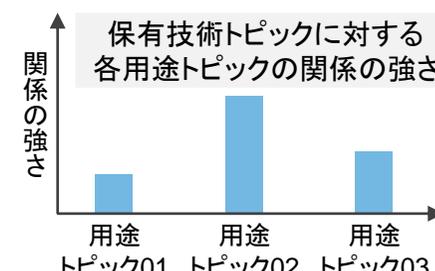
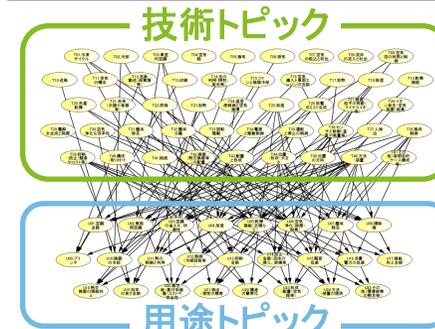
トピックスコアを出願人で集計して、各トピックにおける出願人の特徴を把握する



### C. 用途と技術の関係分析

用途トピックと技術トピックの統計的な関係性をベイジアンネットワークでモデル化する

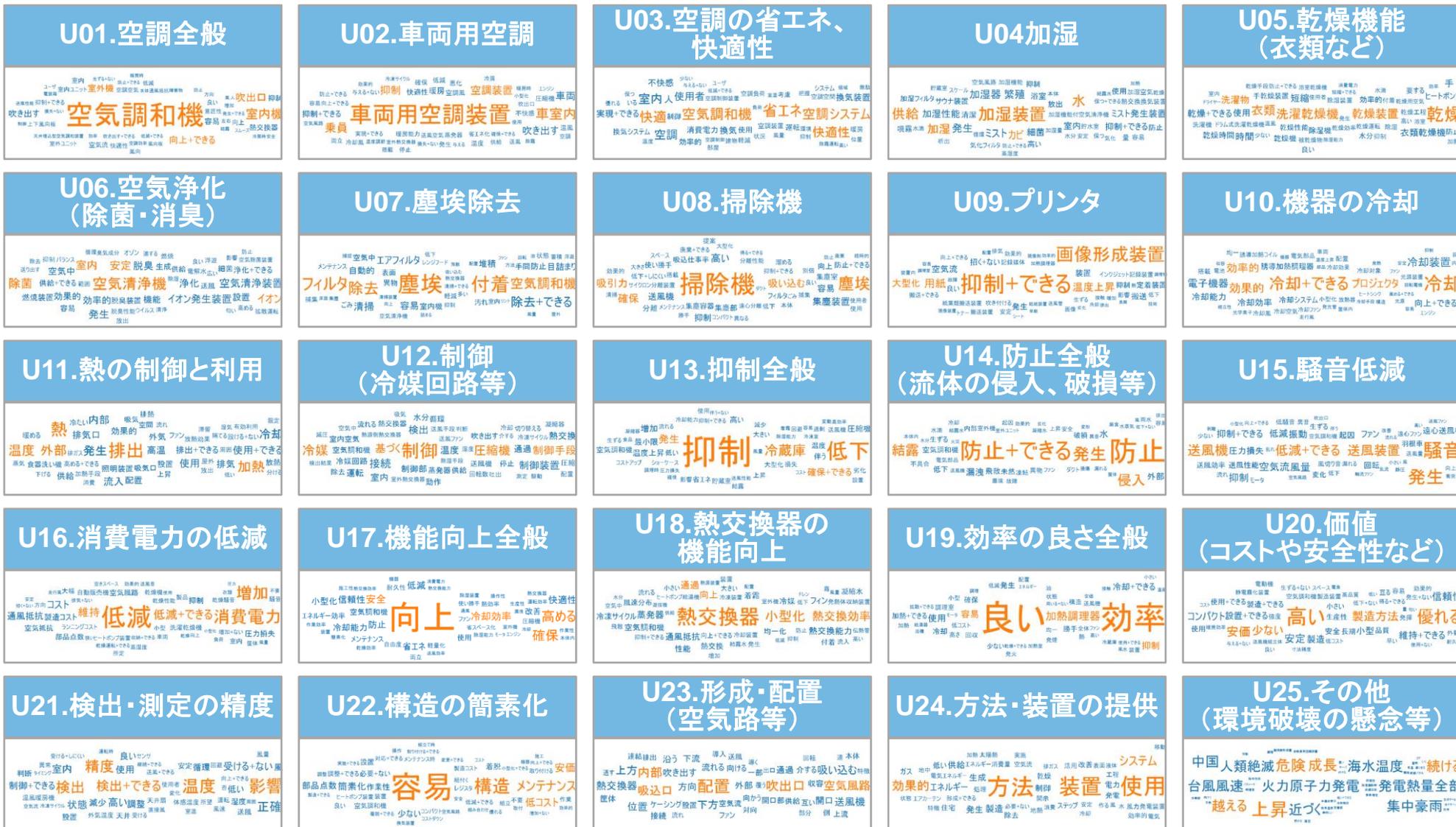
### ベイジアンネットワーク



保有技術と関係のある用途トピックのうち、まだ想定していない用途を探索し、それに関連する元の特許文書を確認することで具体的な新規用途を検討する

# 用途ピック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







# トピックのフラグデータの作成

全特許データに対して各トピックのスコア(該当有無)を計算することで、トピックをベースとした様々な集計・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	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
1	特願2006-XXXX	2006	A社	1	1		0	1	0		0
2	特願2009-XXXX	2009	B社	0	1		1	0	1		0
3	特願2012-XXXX	2012	C社	0	1		1	1	0		0
4	特願2013-XXXX	2013	D社	1	0		0	1	0		1
...	...	...	...	...	...		...	...	...		...
30039	特願2015-XXXX	2015	X社	1	0		1	0	0		1

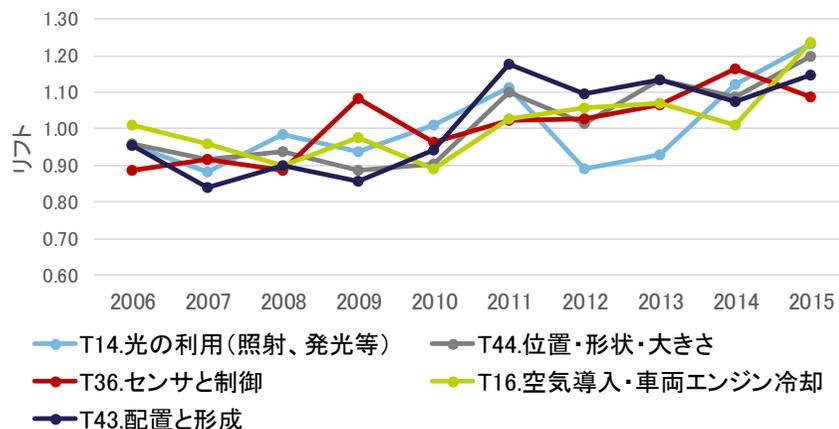
①出願年の集計

②出願人の集計

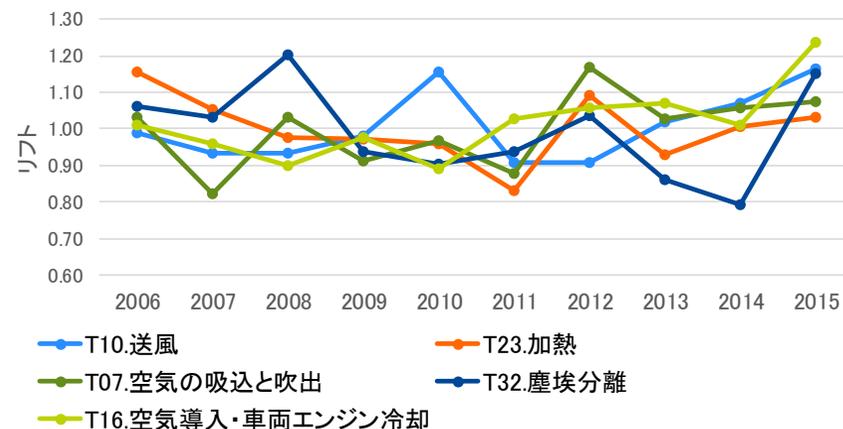
③用途と技術の関連性の分析

短期的には塵埃分離や車両エンジンの冷却に関する技術が、長期的にはプロジェクトなどの光の利用に関する技術が上昇しています

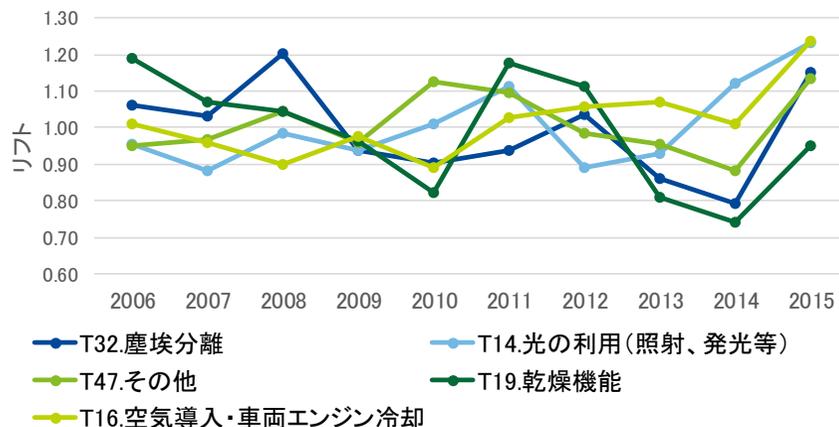
## 2006年からの上昇率 best5



## 2011年からの上昇率 best5



## 2013年からの上昇率 best5



## 集計の仕方

- リフト値を出願年・トピックごとに集計

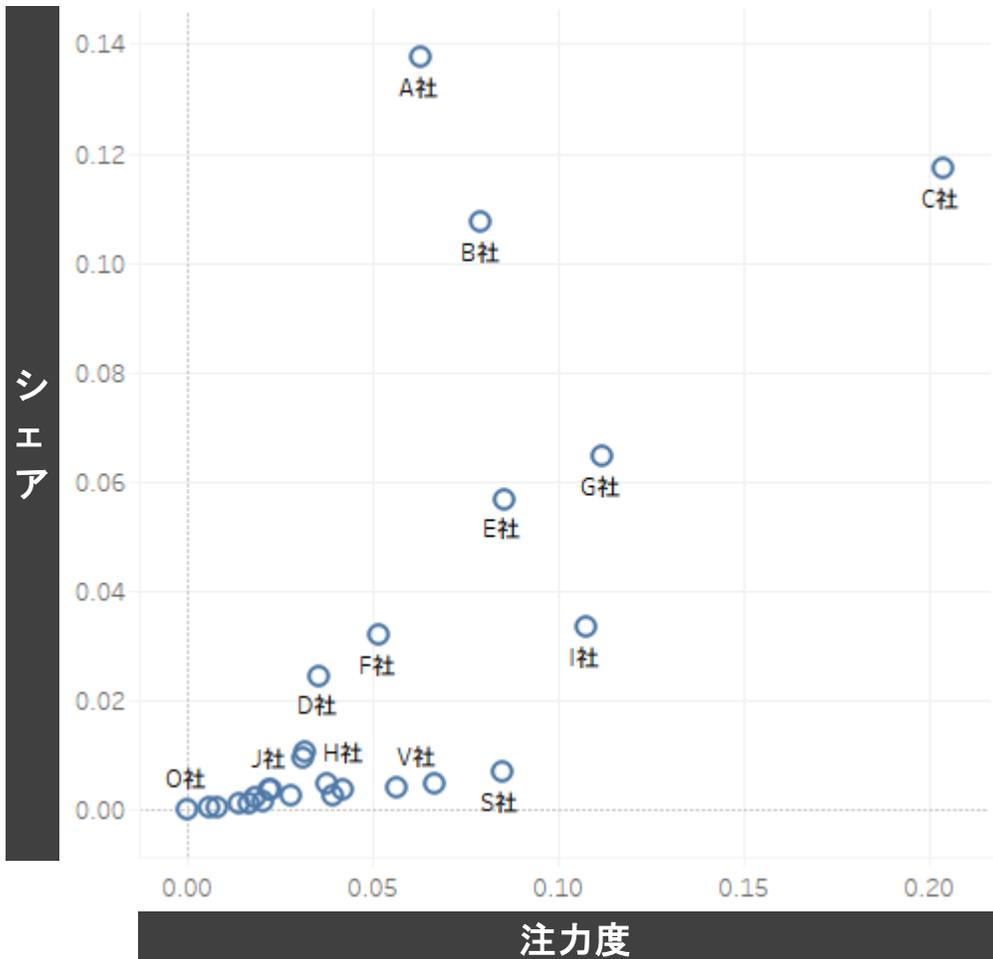
$$P(\text{出願年} | \text{トピック } T_x = 1)$$

$$P(\text{出願年})$$

- その出願年の出願件数割合を平均(=1)として標準化した値

塵埃分離に関する技術は、1社の注力度が高いものの、他にもある程度のシェア・注力度を保有する企業が何社か存在するため、今後連携などの動きも考えられる領域と思われます

## 注力度とシェアの散布図



## 考察と戦略の検討

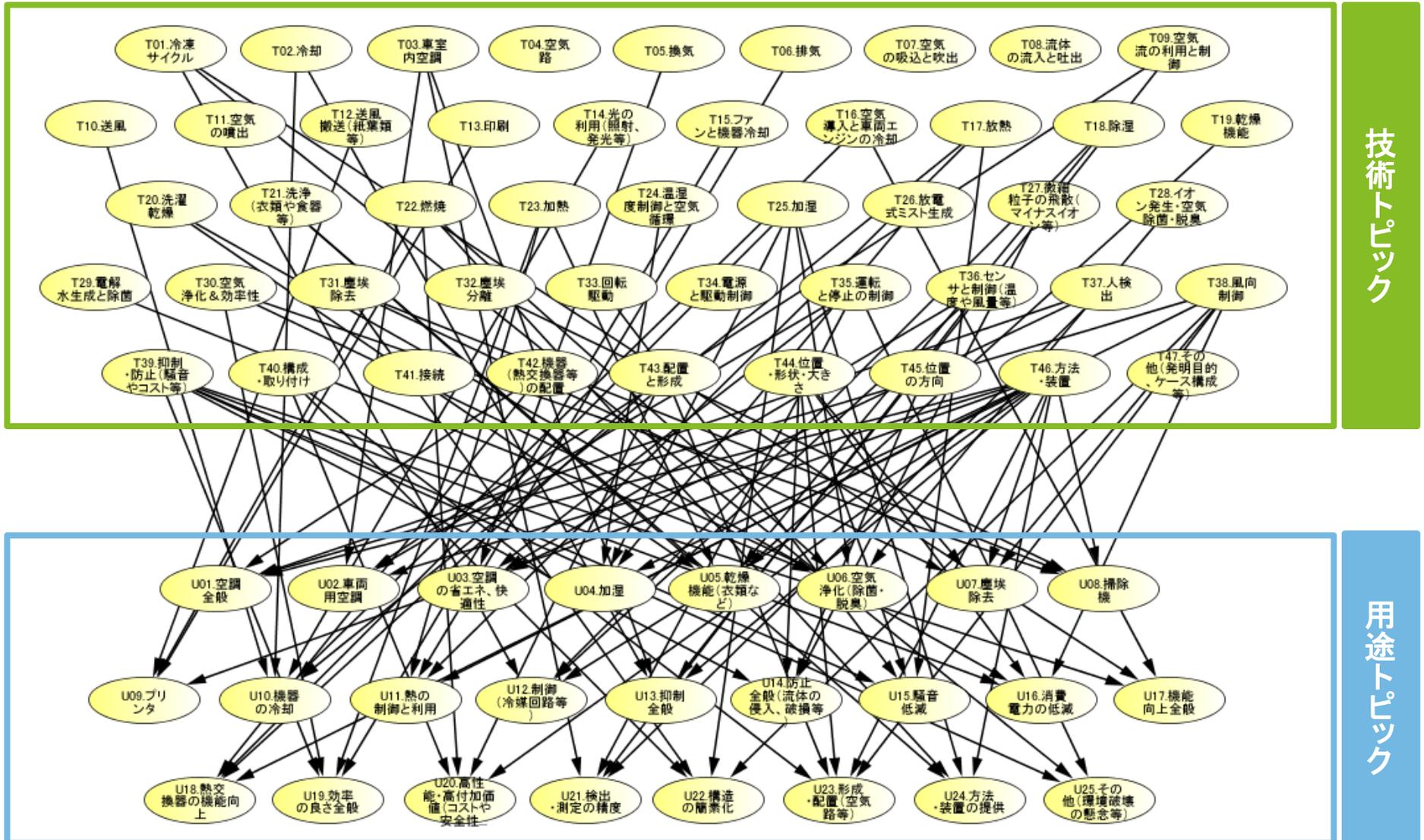
- C社は、高めのシェアを獲得しつつ、他社と比べて注力度がとても高く、高い技術力を保有していると考えられ、今後はよりシェアを伸ばすことで高シェア高注力度のポジションを確立することができると考えられる
- A社とB社は、シェアは高いがまだC社に注力度で劣っているので、例えば規模は中程度だが注力度は比較的高く、技術力があると思われるE社、G社、I社などと連携することで、C社の上のポジションを狙うことができる可能性がある

## 注力度とシェア

- **注力度**:  $P(\text{トピック}T = 1 \mid \text{出願人}X = 1)$ 
  - 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している
- **シェア**:  $P(\text{出願人}X = 1 \mid \text{トピック}T = 1)$ 
  - 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、つまりトピックのなかでどれくらいその出願人が占めているのかを示している

# ベイジアンネットワークを適用した用途と技術の関係分析

用途トピックと技術トピックの{0,1}データにベイジアンネットワークを適用して、技術⇒用途の確率的因果関係をモデル化します

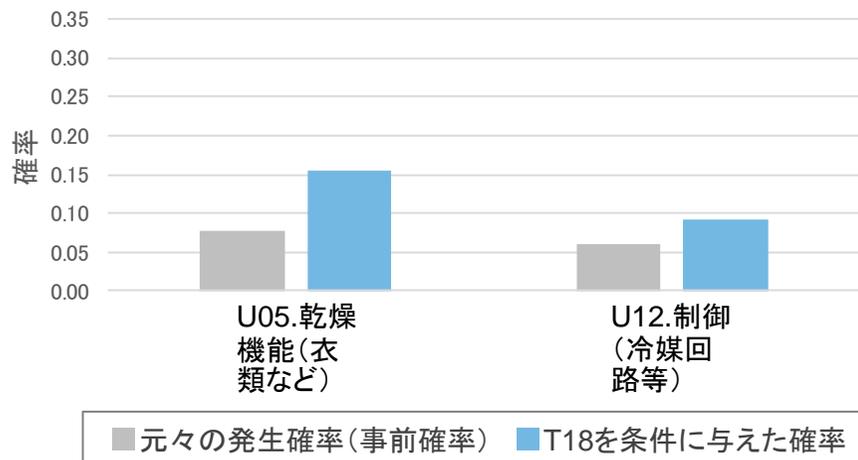


技術トピック

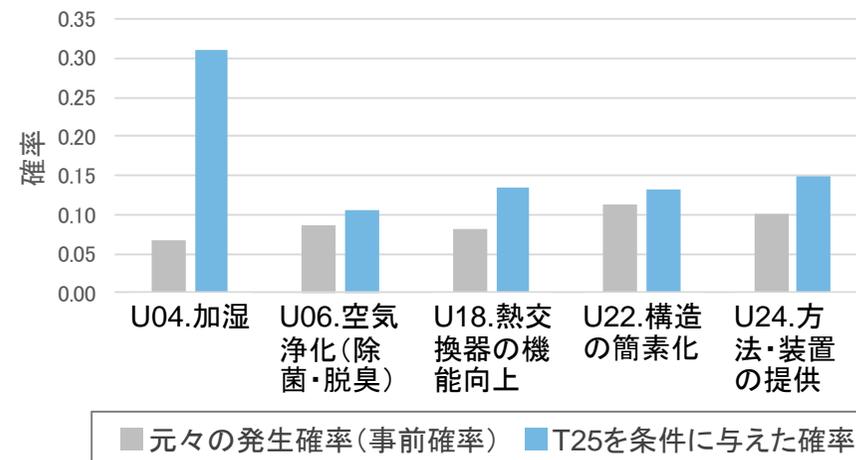
用途トピック

技術トピックを条件に与えたとき、それと確率的因果関係を持つと判定された各用途トピックの確率がどのように変化するのかシミュレーションして、その関連性の強さを確認します

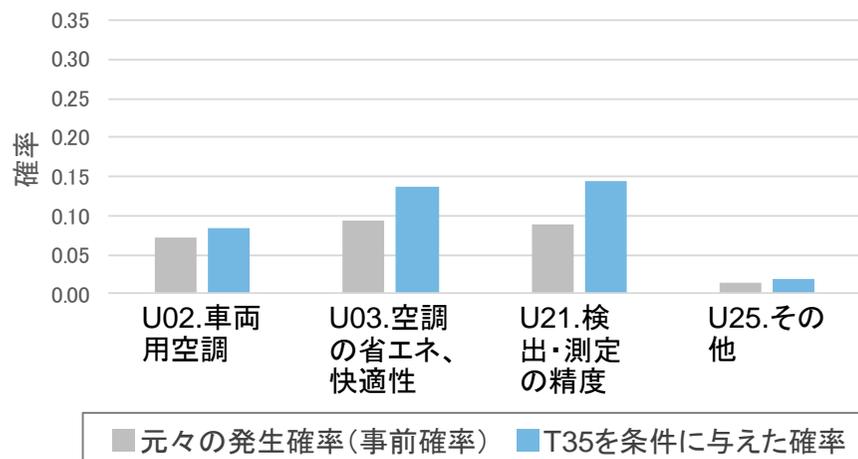
## 「T18.除湿」を条件に与えた結果



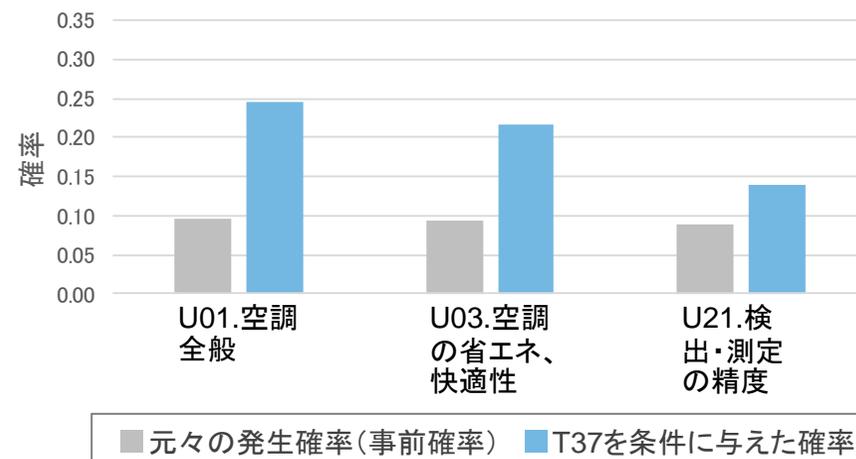
## 「T25.加湿」を条件に与えた結果



## 「T35.運転と停止の制御」を条件に与えた結果

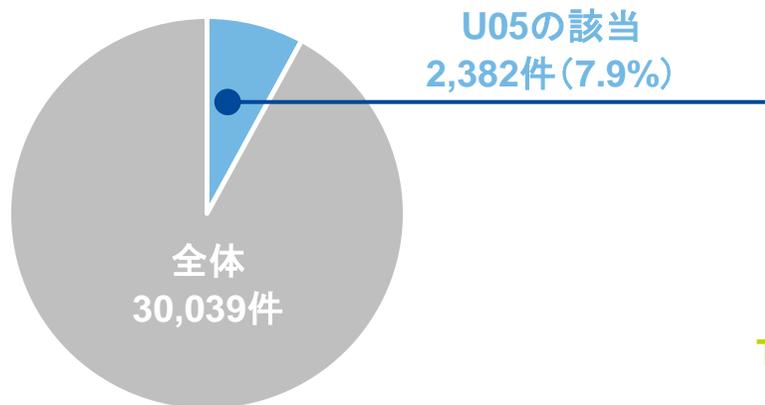


## 「T37.人検出」を条件に与えた結果

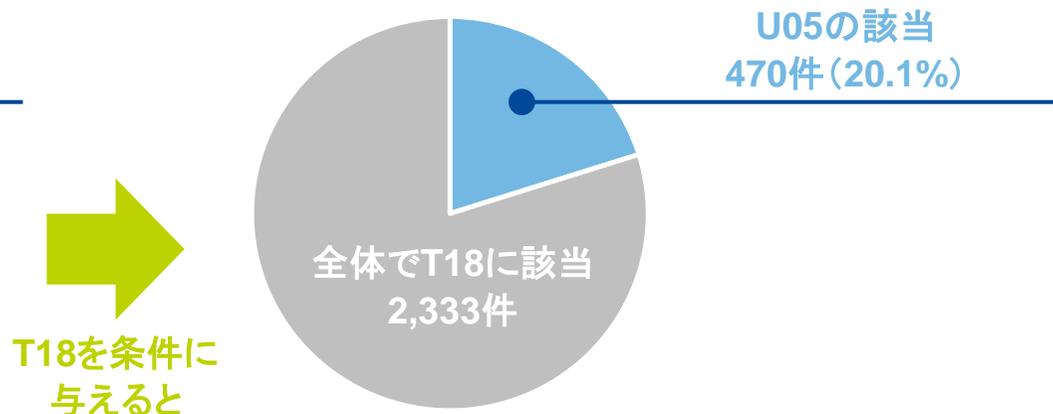


「T18.除湿」の技術の応用先として「U05.乾燥機能」の用途は高い関連性がありますが、出願人Xの保有するT18ではそれがほとんどなく、新規用途となる可能性があります

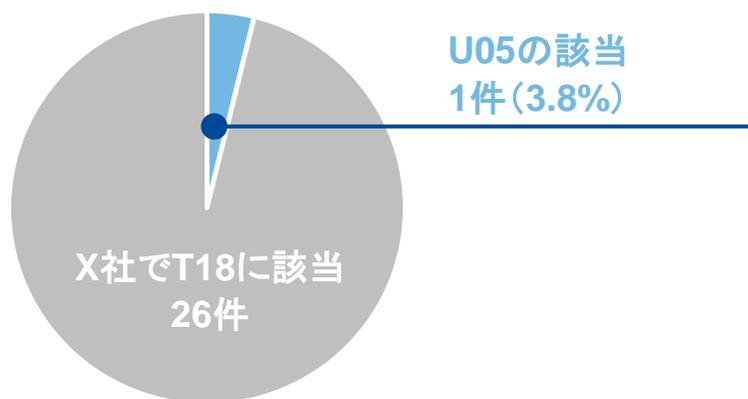
## 全体でのU05の該当割合



## 全体でのT18におけるU05の割合



## 出願人XのT18におけるU05の割合



## 考察

- ベイジアンネットワークのモデルでは、「T18.除湿」に対する「U05.乾燥機能(衣類など)」の関係が見られた
- 全体では、U05の該当は7.9%だが、T18を条件としたときでは、その該当割合が20.1%となり高い関連性が認められる
- しかし、出願人Xでは、T18に該当する特許のうち、U05に該当する特許は1件だけである  
⇒X社の保有するT18はU05での用途も考えられる

## 印刷機の中でインク液を吸収した用紙の湿気をムラなく取り除く乾燥処理技術は、洗濯乾燥機の中で洗濯物をムラなく効率的に乾燥させることにも応用できるかもしれません

### T18がU05で応用されている例

発明の名称
ドラム式洗濯乾燥機
課題
洗濯物を短い時間でムラ無く乾燥させ、乾燥工程の時間を短くすることができるドラム式洗濯乾燥機を提供する。
解決手段
送風機に吸い込まれた空気は、風路切替弁の切り替えにより、ドラム開口部に対向する前側吹出口へ流れたり、回転ドラムの後部に設けられた後側吹出口へ流れたりする。制御装置が風路切替弁の切り替えを制御することによって、恒率乾燥過程時、前側吹出口から乾燥用空気が吹き出し、かつ、減率乾燥過程時、後側吹出口から乾燥用空気が吹き出す。これにより、恒率乾燥過程において乾燥用空気が効果的に当たらなかった、回転ドラムの後端壁側の洗濯物に、乾燥用空気が減率乾燥過程で効果的に当たる。

### 出願人Xの保有するT18の例

発明の名称
インクジェット記録装置及び画像記録方法
課題
処理液の厚みムラを低減するとともに処理液による用紙のコックリングを低減することで、高品質かつ高速の画像記録を可能とするインクジェット記録装置及び画像記録方法を提供する。
解決手段
記録媒体に処理液を付与する処理液付与部の後段には、記録媒体表面に残存する溶媒を蒸発させるプレ加熱部が設けられている。プレ加熱部はIRプレヒータにより記録媒体表面を輻射加熱するとともに、吸引ファンにより記録媒体表面の湿り空気を置換する。液状の処理液が不均一にならないように乾燥処理を施すことで、均一な膜厚を持つ固体状の凝集処理層が形成される。その後、本加熱部による熱風噴射加熱により、コックリング量が所定量以下になるように本加熱処理が施される。

※対外説明用のため要約文は一部加工している

# まとめ

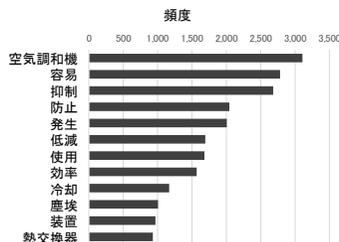
膨大なテキストデータをトピックに変換して解釈を容易にし、テキスト情報内に潜む要因関係をモデル化して、ビジネスアクションに有用な特徴を把握可能にします

# Nomolytics<sup>®</sup>: Narrative Orchestration Modeling Analytics

## テキストマイニング

文章に含まれる単語を抽出し、その出現頻度を集計する

単語抽出



## PLSA 確率的潜在意味解析

単語が出現する文脈を学習し、膨大な単語を複数のトピックにまとめる

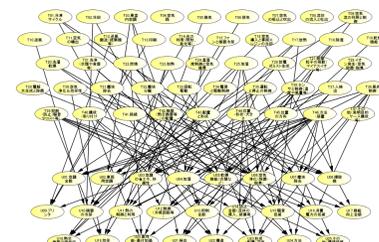
トピック抽出



## 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

モデリング



特許文書に適用することで

特許文書に潜む特徴（トレンドや出願人の動向）をトピックをベースに分かりやすく理解できる

技術と用途の統計的な関係を把握することで、技術の新しい用途のアイデアを創出できる

## Nomolyticsは様々な業務のテキストデータに適用することができます



### 口コミ

- 顧客の関心トピックのターゲット別把握
- 顧客目線での製品や競合の比較分析
- 満足度向上の要因の把握
- 価値観を理解したマーケティング検討



### アンケート

- 自由記述の内容をトピック化
- 自由記述トピックを変数として扱うことで定型設問回答と一緒に分析可能
- 話題を生む要因の把握



### コールセンター履歴

- 問い合わせ内容をトピック化
- 製品別・顧客別の問い合わせ特徴把握
- 問い合わせトピック等の条件から解約確率をシミュレーション



### 特許文書

- 課題と技術のトピックのトレンド把握
- 競合他社の技術動向把握
- 課題と技術のトピックの関係モデル化による保有技術の新規用途探索



### 営業日報

- 営業活動内容のトピック化
- 営業活動トピック等の条件から成約確率をシミュレーション
- 成約要因を把握した効果的な営業教育



### 有価証券報告書

- 各企業の事業内容をトピック化
- 事業トピックとそのトレンド把握
- 各種IR指標と事業トピックの関係分析
- 定性情報からの企業分析、業界分析



### エントリーシート

- 志望動機やPR文のトピック抽出
- 記述内容からの学生の分類・振り分け
- 記述内容と入社後成果の関係分析
- 効率的な人材発掘



### 診療記録

- 診療記録、看護記録のトピック化
- 生活習慣と病状の関係分析
- 治療内容とその経過の関係分析
- 定性情報を用いた効果的な診療支援



### 問題発生レポート

- 不具合やヒヤリハット等のトピック抽出
- 作業環境等の条件から問題の発生確率をシミュレーション
- 効果的な製品や作業環境の改善支援

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](mailto: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

