



Analytics Design Lab

VOCデータの分析サービス

株式会社アナリティクスデザインラボ

1. 会社概要・サービス概要	3
2. テキストマイニング分析の課題と解決	6
3. 当社の独自技術「Nomolytics」のご紹介	13
4. NomolyticsによるVOC分析サービスの提供例	19

1. 会社概要・サービス概要

企業様のデータ分析・活用を支援させていただくコンサルティング会社です

会社概要：株式会社アナリティクスデザインラボ

社名	株式会社アナリティクスデザインラボ Analytics Design Lab Inc.
所在地	〒164-0002 東京都中野区上高田1-2-51-303
設立	2017年6月1日
資本金	5,000,000円
代表取締役	野守耕爾
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
取引銀行	三菱東京UFJ銀行 東中野支店
導入ツール	・Text Mining Studio (NTTデータ数理システム) ・Visual Mining Studio (NTTデータ数理システム) ・BayoLink (NTTデータ数理システム) ・SPSS Modeler (IBM) ・JMP (SAS) ・Tableau Desktop (Tableau Software)

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

お客様が実施されるデータ分析に対して当社が助言する「顧問アドバイザーサービス」と、当社がデータ分析を実行しご報告する「分析プロジェクトサービス」をご用意しています

顧問アドバイザーサービス

お客様が実施されるデータ分析・活用のお取り組みについて、専門的な観点から助言いたします

分析プロジェクトサービス

お客様の課題に応じて、データの分析と活用のプロセスを当社が設計・実行し、報告いたします

サービスの比較

お客様	分析実施者	当社
会議・メール形式での助言	提供形式	会議形式での実施報告
なし	納品成果物	報告書
3ヶ月単位	契約期間	プロジェクトの内容による

2. テキストマイニング分析の課題と解決

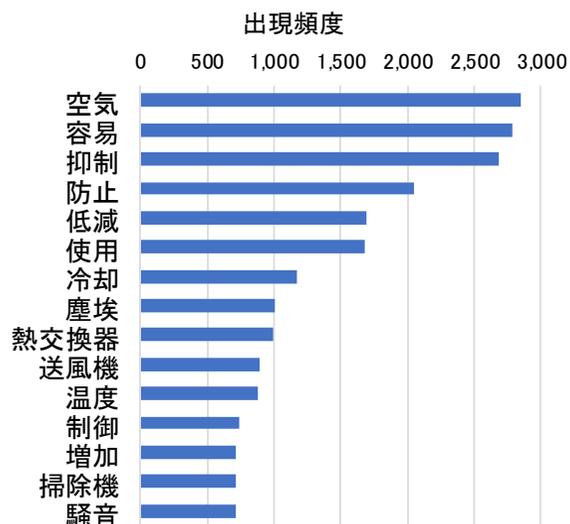
テキストマイニングでよくある可視化のアウトプット

テキストマイニングの分析では、文章に含まれる単語や係り受け表現をベースとした集計・統計分析を実行することで、文章の特徴を可視化して全体像を把握します

テキストマイニングの可視化例

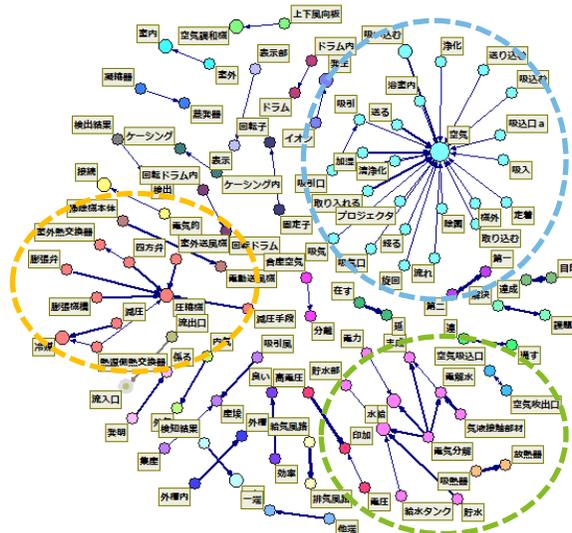
頻度集計

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



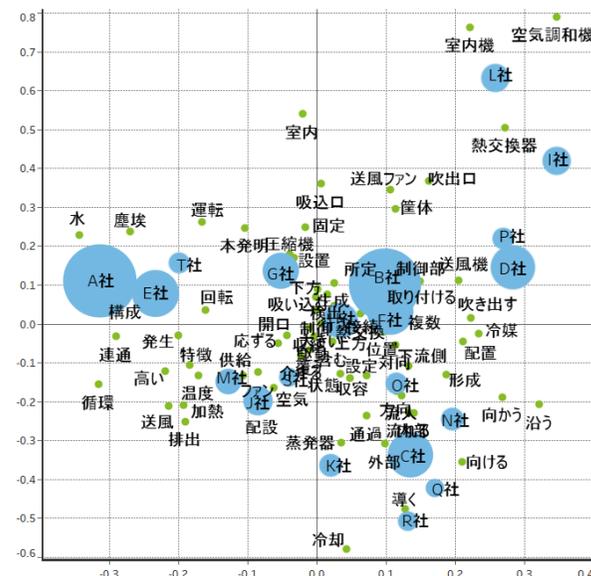
共起ネットワーク

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



コレスポネンス分析

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



※特許データをテキストマイニングした例を掲載

テキストマイニングの課題とその解決技術

テキストマイニングの分析は単語ベースの結果で解釈が難しく、特徴がモデル化されていないので状況の変化に応じたシミュレーションができませんが、これをAI技術で解決します

課題① 解釈がしづらい

- ◆ 単語をベースにした特徴の可視化は複雑であり、結果の解釈が困難になる
- ◆ 単語を人がある程度グルーピングして分析することもあるが、その作業は主観的で負荷が大きい

解決技術

単語をクラスタリングする
AI技術

PLSA
確率的潜在意味解析

- ◆ 単語の出現状況を学習することで、使われ方の似ている単語をまとめ上げ、トピックを抽出する
- ◆ 膨大な「単語」ではなく、いくつかの「トピック」をベースに特徴を可視化し、解釈を容易にする

課題② 現状把握に留まる

- ◆ 記述傾向の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
- ◆ 現状から状況が変化したときに、それに伴って結果がどう影響を受けるのかシミュレーションできない

解決技術

現象をモデリングする
AI技術

ベイジアンネットワーク

- ◆ PLSAで抽出されたトピックや、その他属性情報などの要因間に潜む要因関係をモデル化する
- ◆ 各要因の条件を変化させると他の要因がどのような挙動をとるのか確率的にシミュレーションする

PLSAは、トピックモデルと呼ばれるAI技術で、テキストマイニングで抽出された大量の単語をいくつかのトピックに集約して類型化することができます

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して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

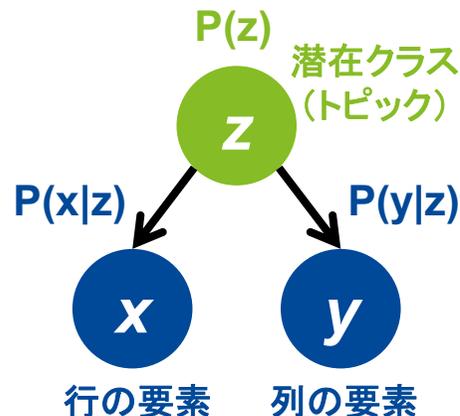
「文書×単語」行列 (共起行列)

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	2		0	1
...						

文書ID	トピック 1	トピック 2	...	トピック 15
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(x|z), P(y|z), P(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

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

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

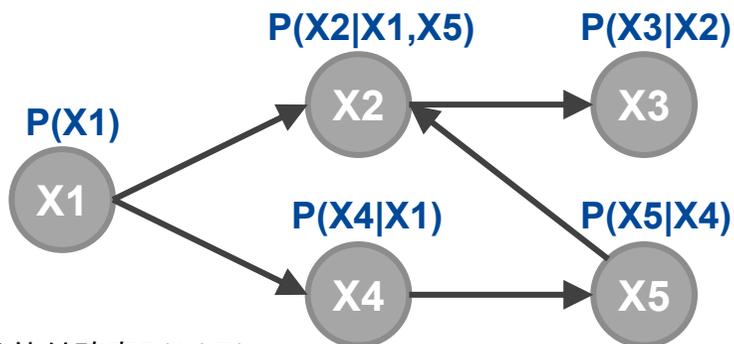
ソフトクラスタリングできる

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

ベイジアンネットワークは、ベイズ推論に基づいたAI技術で、変数と変数の間に潜む確率的な因果関係をネットワーク構造で探索し、指定条件下の確率をシミュレーションできます

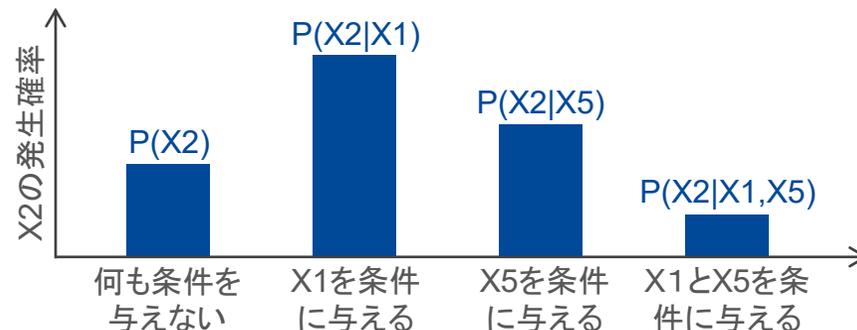
ベイジアンネットワークの概要

- 複数の変数の確率的な因果関係をネットワーク構造で表わし、ある変数の状態を条件として与えたときの他の変数の条件付確率を推論することができる
- 目的変数と説明変数の区別はなく、様々な方向から変数の確率シミュレーションができる
- 全ての変数は質的変数(カテゴリカル変数)となるため、量的変数の場合は閾値を設けてカテゴリに分割する
- 確率論の非線形処理によるモデル化のため、非線形の関係や相互作用が生じる現象でも記述できる



※条件付確率 $P(A|B)$
事象Bが起こる条件の下で事象Aの起こる確率

確率的因果関係と相互作用



- X2の発生確率は、何も条件を与えない時(事前確率)と比べて、X1やX5を条件に与えると確率が上昇する
⇒X1やX5はX2の発生に関して"確率的な"因果関係がある
- しかし、X1とX5の両方を条件に与えると、元々の事前確率よりも確率が下がってしまう
⇒X1とX5はX2に対して相互作用がある(X1とX5は相性が悪い)

ベイジアンネットワークのメリット

現象を理解して柔軟にシミュレーションできる

目的変数、説明変数の区別なく変数の関係をモデル化するので、現象の構造を理解でき、推論変数と条件変数を自由に指定して確率推論できる

効果を発揮する有用な条件を発見できる

ある条件のときにだけ効果が現れるといった相互作用がある場合でも、確率的に意味のある関係としてモデル化することができる

生成AIは、文脈を理解でき、汎用性が高く扱い易いですが、正確性や再現性で課題があり、テキストマイニングは、時間を要し、スキルも求められますが、正確な分析が可能です

生成系AI

テキストマイニング

メリット



- 言葉による柔軟な指示を出すことができ、タスクの**汎用性**が極めて高い
- アウトプットまでが**高速**である
- **文脈**を捉えた分析ができる
- 専門知識がなくても**扱いやすい**

- 単語の出現頻度という事実ベースの分析で、**正確で客観的**な分析ができる
- 分析は**再現可能**で、分析のプロセスもその結果も根拠が明確である
- **定量的**な集計や統計分析に優れている
- **大量のデータ**でも分析できる

デメリット



- 事前学習した汎用モデルによる推論ベースのアウトプットであり、**ハルシネーション**が起きるリスクが常にあり、特に**定量的な情報**の処理は正確でないことが多い
- 生成プロセスが**ブラックボックス**となる
- 同じプロンプトで回答が異なり**再現性がない**
- **大量のデータ**を用いた分析は難しい

- 単語の出現頻度だけに基づく単純な分析であり、**文脈の理解には限界**がある
- 分析に多くの**時間を要し**、専門知識や経験も求められる
- 分析の質が**分析者のスキルに依存**する

人が理解し考えることで生まれるデータ分析の価値

最新のAIならデータ分析の自動化も夢ではなさそうですが、現場で知見ある人が自ら分析するからこそ、深い洞察が得られ、実践的で納得感のあるアクションにつながると考えます



3. 当社の独自技術「Nomolytics」のご紹介

テキストマイニング × PLSA × ベイジアンネットワーク

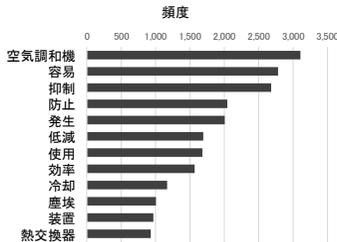
膨大なテキストデータをトピックに変換して解釈を容易にし、テキスト情報内に潜む要因関係をモデル化して、ビジネスアクションに有用な特徴を把握可能にします

Nomolytics : 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

単語が出現する特徴を学習し、膨大な単語を複数のトピックにまとめる

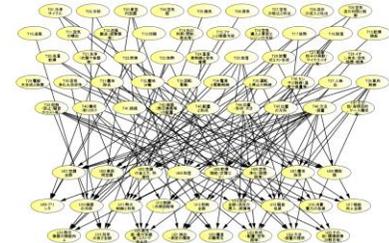
トピック化



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

因果分析



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

Nomolyticsを構成する各分析手法はNTTデータ数理システム社の分析ソフトウェアを使用することで実行できます

Nomolytics : Narrative Orchestration Modeling Analytics

テキストマイニング

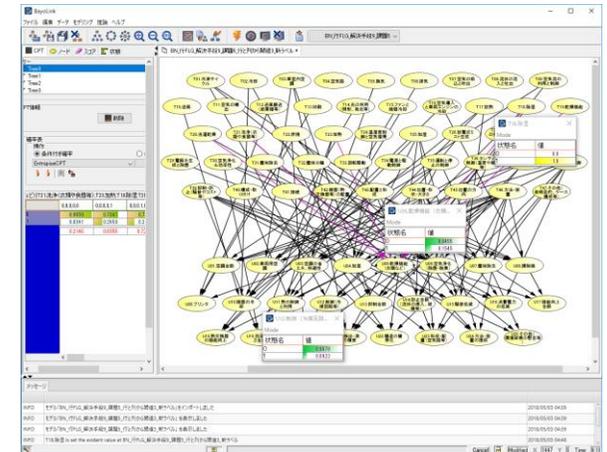
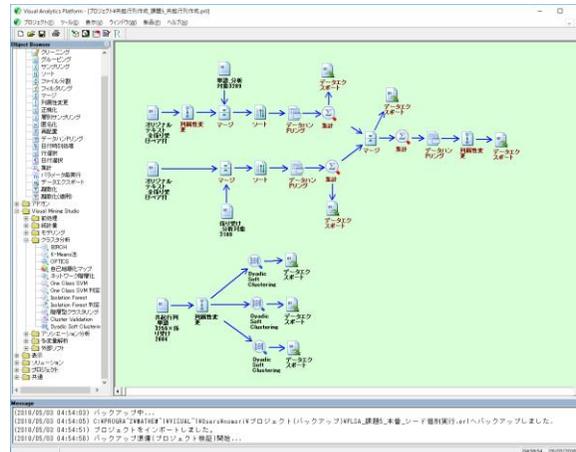
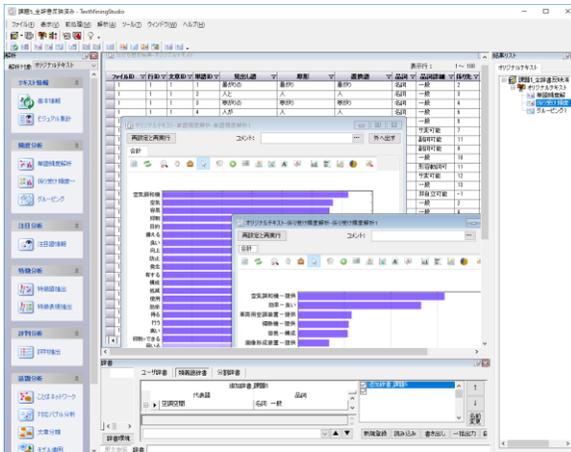
PLSA
確率的潜在意味解析

ベイジアンネットワーク

Text Mining
Studio

Visual Mining
Studio

BayoLink



Nomolyticsは様々なテキストデータに適用できますが、口コミ、アンケート、コールセンター等VOCの分析では、顧客の声をトピックに類型化し、顧客満足の要因を発見できます

口コミ

- 顧客ターゲット別の関心事トピックを把握
- 製品・サービス別のトピックを把握
- 口コミ得点に寄与するトピックを把握
- ニーズに応じたマーケティングを検討

アンケート

- 自由記述回答の内容をトピックで把握
- トピック化された自由記述回答と通常の定型設問回答の関係を統計分析
- 顧客満足度に寄与するトピックを把握

コールセンター履歴

- 問い合わせ内容をトピックで把握
- 製品別・顧客別のトピック傾向を把握
- 解約・退会に寄与するトピックを把握
- 満足度向上、顧客離反抑制の施策検討

特許文書

- 特許文書の内容をトピックで把握
- トレンドや競合他社の動向を把握
- 用途と技術の関係分析から用途実現の技術戦略や保有技術の新規用途を検討

営業日報

- 営業活動内容をトピックで把握
- 営業属性別のトピック傾向を把握
- 成約に寄与するトピックを把握
- 成約のための効果的な営業教育を検討

有価証券報告書

- 企業・業界の事業内容をトピックで把握
- 事業内容トピックのトレンドを把握
- 好業績に寄与する事業トピックを把握
- 定性情報から行う企業分析・業界分析

エントリーシート

- 志望動機やPR文のトピックを把握
- 記述トピックに基づいて学生を分類
- トピック傾向から面接の質問内容を検討
- 選考通過に寄与するトピックを把握

診療・看護記録

- 診療記録、看護記録をトピックで把握
- 患者の属性別のトピック傾向を把握
- 検査指標に寄与する定性情報を把握
- 定性情報も用いた診療・助言を検討

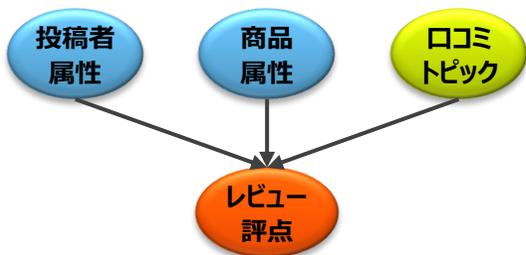
問題発生レポート

- 不具合やヒヤリハットをトピックで整理
- 作業環境別のトピック傾向を把握
- 重大問題に寄与するトピックを把握
- 問題を抑制する作業・環境改善を検討

VOCデータにNomolyticsを適用することで、VOCのトピックを軸に、顧客の行動の要因構造をベイジアンネットワークでモデル化でき、そのキードライバーを定量的に把握できます

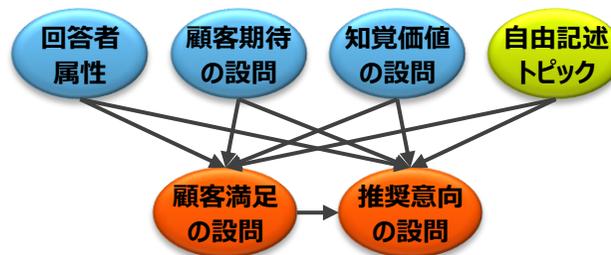
口コミ

口コミのレビュー評点の要因として、投稿者の属性や商品の属性、口コミトピックの効果把握する



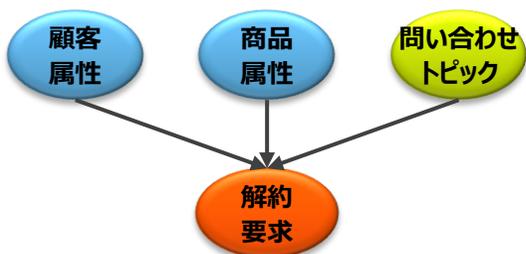
アンケート

顧客満足度の要因として、回答者の属性や選択式設問の回答結果、自由記述トピックの効果把握する



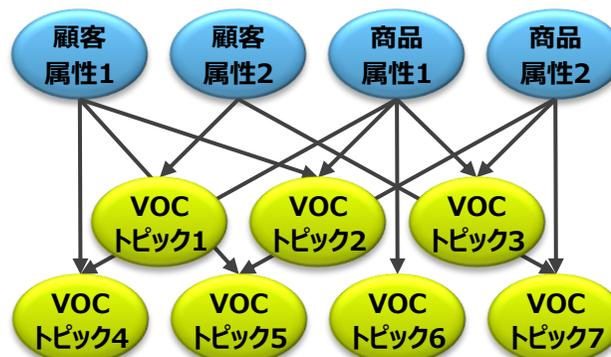
コールセンター問い合わせ履歴

解約などの要求行動の要因として、顧客の属性や商品の属性、問い合わせトピックの効果把握する



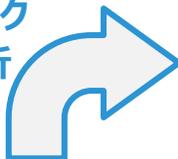
VOC共通

どのような顧客の属性や商品の属性で、どのような顧客の声が寄せられるか把握する



NomolyticsでVOCデータを分析することで、そのコメント内容をトピック化して各属性の傾向や要因関係を分析でき、顧客目線の業務改善やマーケティング戦略を検討できます

属性×トピック
の特徴分析

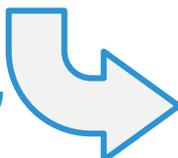


A VOCのトピック化

VOCのテキストデータにテキストマイニング×PLSA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



ベイジアンネットワーク
による要因関係分析



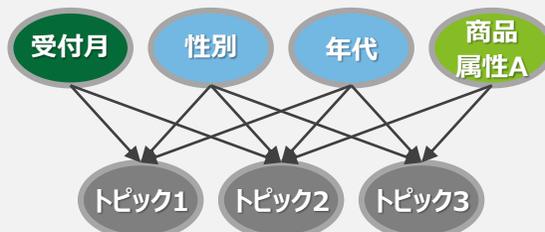
B トピック×属性の傾向分析

顧客属性別、商品属性別に各トピックの関連度を計算することで、各属性におけるコメントの傾向を把握します

	トピック1	トピック2	トピック3	トピック4
20代30代	0.6	1.2	4.3	0.9
40代50代	2.4	1.5	1.6	2.3
60代70代	3.8	1.1	0.4	1.2
商品属性1	0.5	3.4	1.2	2.2
商品属性2	4.1	0.8	0.3	1.0

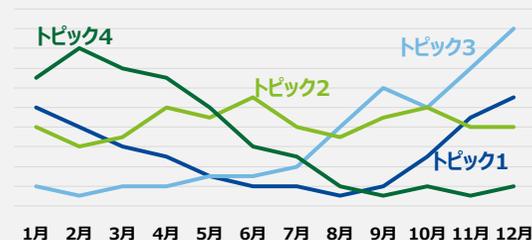
D トピックの要因関係の分析

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



C トピックのトレンド分析

時系列情報と各トピックの関連度を計算することで、トピックのトレンドや季節性を把握します



E 行動の要因と確率シミュレーション

データに満足度や解約などの行動情報があれば、その行動に影響を与える要因をトピックや属性から探索し、行動を確率的に予測したり、行動を制御する条件を把握します



4. NomolyticsによるVOC分析サービスの提供例

分析案(B). トピック×属性の傾向分析

データの属性情報(顧客属性や商品属性)を軸にトピックのスコアを集計することで、各属性別におけるトピックの傾向を定量的に可視化します

目的

- VOCデータに紐づく属性情報(顧客属性や商品属性)を軸に、抽出したトピックのスコアデータをクロス集計することで、各属性別のトピックの傾向を定量的に把握する

実施方法

- (B-1)属性データの前処理
VOCデータに紐づく属性情報(顧客や商品など)のカテゴリ分類を整理する。
- (B-2)集計用データの作成
トピックのスコアの閾値を設定することで、トピックの該当有無のフラグデータを作成し、これをステップB以降の集計用データセットとする。
- (B-3)トピック×属性のクロス集計
各属性の軸でトピックのスコアデータを集計・可視化することで、属性カテゴリ別のVOCの傾向を定量的に把握する。
- (B-4)トピック×属性の特徴表現の抽出
(B-3)集計結果において、強い傾向を示す対象に該当するテキストデータに再度テキストマイニングを実行して特徴キーワードを抽出する。
- (B-5)トピック×属性の特徴文章の抽出
(B-3)集計結果において、強い傾向を示す対象に該当する元の文章を確認することで、属性×トピックの傾向の詳細を把握する。

実施期間の目安

- データ件数が1~3万件程度の場合: 約2~3週間
- データ件数が10~30万件程度の場合: 約3~4週間

分析の効果

- 顧客属性とトピックの関係から、どのような顧客はどのようなトピックに関心があるのか定量的に把握できる
- 商品属性とトピックの関係から、どのような商品はどのような意見・問い合わせが多いのか定量的に把握できる
- 各顧客属性の関心トピックとそのトピックの特徴を示す商品属性に基づき、誰にどんな商品を提供するとトピックの観点からマッチするののかという効果的なマーケティングを検討できる

アウトプットイメージ

VOCデータ

データID	性別	年代	トピック1	トピック2	トピック3	商品A	商品B	...
1	男性	40代	3.1	1.5	0.9	属性1	属性3	
2	女性	20代	0.5	4.6	2.0	属性2	属性1	
3	女性	30代	4.8	0.7	6.7	属性4	属性2	
...								

顧客×トピックの集計

商品×トピックの集計

	トピック1	トピック2	トピック3	トピック4
20代30代	0.6	1.2	4.3	0.9
40代50代	2.4	1.5	1.6	2.3
60代70代	3.8	1.1	0.4	1.2
商品属性1	0.5	3.4	1.2	2.2
商品属性2	4.1	0.8	0.3	1.0

分析案(C). トピックのトレンド分析

データの時系列情報を軸にトピックのスコアを集計することで、各トピックのトレンドを可視化します

目的

- VOCデータに紐づく時系列情報を軸に、抽出したトピックのスコアデータをクロス集計することで、各トピックのトレンド把握する

実施方法

- (C-1)トピック×時間情報のクロス集計
時間情報(問い合わせの受付年月など)の軸でトピックのスコアデータを集計することで、トピックの出現のトレンドを可視化する。
- (C-2)傾向の詳細分析
(C-1)の集計結果において、強い傾向を示す対象にはそれに該当するテキストデータに再度テキストマイニングを実行して特徴キーワードを抽出したり、その元の文章を確認することで、傾向の詳細を把握する。

実施期間の目安

- データ件数が1~3万件程度の場合: 約0.5週間
- データ件数が10~30万件程度の場合: 約1週間

分析の効果

- 最近増加している問い合わせトピックを把握することで優先的に対策を講じるべき対象を把握できる
- 季節によって増加する問い合わせトピックの傾向を把握することで、それが増加する前に対策を検討できる
- 時間帯によって変動する問い合わせトピックの傾向を把握することで、時間帯別のオペレータの配置を検討できる
- 何か施策を実施した前後での各問い合わせトピックの増減を把握することで、その施策の効果を検証できる

アウトプットイメージ

VOCデータ

データID	受付年	受付月	トピック1	トピック2	トピック3	...
1	2019年	11月	3.1	1.5	0.9	
2	2019年	12月	0.5	4.6	2.0	
3	2020年	1月	4.8	0.7	6.7	
...						

時系列×トピックの集計



分析案(D). トピックの要因関係の分析

トピックと属性との関係構造を確率的因果モデルのAI技術(ベイジアンネットワーク)でモデル化し、トピックの発生確率を各属性の複合条件からシミュレーションします

目的

- トピックのスコアデータを用いて、トピックに対する各属性情報(顧客属性、商品属性、時間属性など)の確率的因果関係をベイジアンネットワークでモデル化することで、各トピックの発生に寄与する複合要因を把握し、どのような顧客にどのような商品が渡るとどのようなトピックが発生しやすいのか確率的にシミュレーションする

実施方法

- (D-1) 注目トピックの要因と効果の分析**
注目トピックに対する各属性の確率的因果関係をベイジアンネットワークでモデル化し、各トピックの発生にはどの属性が複合的な要因となっているか把握する。モデルを用いて、各属性情報に条件に与えたときの各トピックの発生確率をシミュレーションすることで、どのような顧客にどのような商品が渡るとどのようなトピックが発生しやすいのか把握する。また各トピックの発生確率を最も高くするあるいは最も低くする効果的な要因条件を把握する。
- (D-2) 注目トピックに影響を与える要因の特徴文章の抽出**
(D-1)の結果において、注目トピックの発生確率に影響を与える要因に該当する元の文章を確認することで、それらの要因と効果の詳細を把握する。

実施期間の目安

- データ件数が1~3万件程度の場合: 約2週間
- データ件数が10~30万件程度の場合: 約3週間

分析の効果

- 新商品を市場に投入する前に、その商品に該当する属性とターゲット層を条件にセットし、その条件下での各トピックの発生確率をモデルでシミュレーションすることで、その新商品はどのような問い合わせが起こり得るのか事前に評価できる
- あるトピックの発生確率を最大化するような、あるいは最小化するような顧客属性と商品属性の組み合わせを計算することで、話題にされるトピックを狙った新商品とターゲットの効果的な組み合わせを企画できる

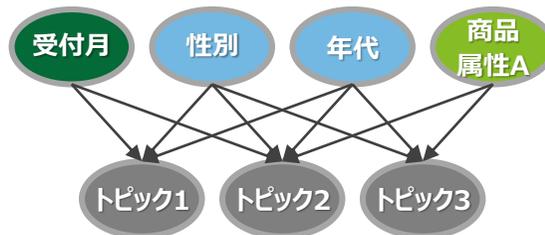
アウトプットイメージ

VOCデータ

データID	受付月	性別	年代	商品A	トピック1	トピック2	トピック3	...
1	11月	男性	40代	属性1	3.1	1.5	0.9	
2	12月	女性	20代	属性2	0.5	4.6	2.0	
3	1月	女性	30代	属性4	4.8	0.7	6.7	
...								

illustrative

複数の属性×トピックの確率的因果関係モデル



分析案(E). 評価の要因関係の分析

満足度などの評価に影響を与える要因(トピックや属性)との関係構造をベイジアンネットワークでモデル化し、各条件からの評価予測や評価を高める条件を把握します

目的

- 満足度などの評価情報がある場合、トピックのスコアデータを用いて、評価情報に対する各トピック及び属性情報(顧客属性や商品属性)の確率的因果関係をベイジアンネットワークでモデル化することで、評価に影響を与える要因を把握し、各要因条件から評価の確率をシミュレーションしたり、評価を向上させる効果的な要因条件を把握する

分析の効果

- 新商品を市場に投入する前に、その商品に該当する属性とターゲット層、さらに想定されるVOCのトピックを条件にセットし、その条件下での評価の確率をモデルでシミュレーションすることで、その条件における評価の程度を予測できる
- 評価をより高めるような効果的な条件を把握し、顧客目線の満足度の高い新しい価値の創造を検討できる

実施方法

- (E-1) 評価の要因と効果の分析
評価情報に対する各トピック及び属性の確率的因果関係をベイジアンネットワークでモデル化し、評価に影響を与える要因を把握する。モデルを用いて、各要因(トピック及び属性)に条件に与えたときの評価の確率をシミュレーションすることで、どのような顧客にどのような商品が渡り、どのようなトピックを話題にされると評価にどれくらいの影響を与えるのか把握する。また評価を向上させる効果的な要因条件も把握する。
- (E-2) 評価に影響を与える要因の特徴文章の抽出
(E-1)の結果において、評価情報の発生確率に影響を与える要因に該当する元の文章を確認することで、それらの要因と効果の詳細を把握する。

アウトプットイメージ

VOCデータ

データID	満足度	性別	年代	商品A	商品B	トピック1	トピック2	トピック3	...
1	2	男性	40代	属性1	属性3	3.1	1.5	0.9	
2	5	女性	20代	属性2	属性1	0.5	4.6	2.0	
3	4	女性	30代	属性4	属性2	4.8	0.7	6.7	
...									

illustrative

評価の確率的因果関係モデル



実施期間の目安

- データ件数が1~3万件程度の場合: 約1~2週間
- データ件数が10~30万件程度の場合: 約1.~2週間

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。

<https://www.analyticsdlab.co.jp/>

株式会社アナリティクスデザインラボ

