


 Fact Sheet

主な機能

SAS Contextual Analysisを使用すると、文書を分類して事実を抽出し、文書毎のセンチメント(評判)を深く理解するまでの作業を一貫した操作環境で実行し、テキストデータから短時間で洞察を導き出すことができます。また、学習用コーパスを作成しなくてもルールを定義できます。

ビジネスメリット

大量のテキストデータから新たな課題の芽を見つけ出したり、傾向や動向を見極めたりする作業が容易になります。得られた情報はビジネスや顧客の理解を深めるために活用できるほか、業界規制を遵守しながら競争優位性を維持する目的にも役立ちます。

対象ユーザー

SAS Contextual Analysisは、大量の文書に埋もれがちで、かつ、手作業で文書の整理や内容の確認を行う時間や人的リソースのない企業や組織での利用を想定しています。短期間で本稼働を開始でき、カスタマイズにも対応しているため、IT部門ユーザーと業務部門ユーザーの双方にとって理想的なツールです。

SAS® Contextual Analysis

非構造化テキストを構造化し、より多くの洞察を組織全体に提供

ビジネス文書や顧客とのやり取りのメールなど、さまざまな非構造化データは次々と生成されていくため、手作業で整理・確認するのは現実的ではない大量の資料があつという間に蓄積されます。しかし、適切な技術を導入すると、こうした情報の蓄積を宝の山に変えることができます。

SAS Contextual Analysisは、非構造化テキストからビジネスに関する貴重な洞察を導き出すために役立つ高度な言語処理ソリューションです。機械学習テクノロジーと各分野の専門知識を組み合わせ、データに潜んでいる傾向やテーマを浮き彫りにすることができます。そこから得られる洞察は他の方法では決して明らかにならない可能性が高く、新たな視点からビジネスの理解を深めるために役立ちます。

SAS Contextual Analysisを支える技術は高度ですが、使い方は難しくありません。タクソノミー(分類法)のための学習用コーパスをユーザーが作成しなくても、出発点として利用できるトピックが自動的に検出されます。また、柔軟性も優れています。機械学習によって生成されたカテゴリー・ルールをより具体的なものに改良し、より多くのビジネス価値を生み出すために欠かせない正確な洞察を導き出すことが可能です。

主な利点

- **テキスト分析に関連した全ての作業を単一の統合システムで容易に管理**

SAS Contextual Analysisは短期間で本稼働を開始でき、必要に応じてルールをカスタマイズすることもできる統合環境です。すべての作業をこの環境内で行えます。また、高度な自動化により、ユーザーはソフトウェアの導入後すぐに探索的な作業を開始できるほか、自然言語処理技術を用いてユーザー独自の定義を設定することも可能です。

- **ガイダンス付きのハイブリッド・アプローチによるコントロール性の強化**

SAS Contextual Analysisでは、テキスト分析をカスタマイズする際にハイブリッド型のアプローチを利用できます。これは、機械学習で自動的に獲得した洞察に各分野の専門知識を組み合わせる手法です。分析エンジンは、この統合データに基づいて構文を解析し、共通テーマ、カテゴリー・ルール、ドキュメント・センチメントを明らかにします。また、ユーザーはいつでも機械学習の結果を確認し、特定のニーズに合わせて機械学習の結果を調整することができます。つまり、ハイブリッド・アプローチの採用により、柔軟性、コントロール性、有効性が高まり、より短期間で目的に適した結果が得られます。

- **さらに広範な用途の基礎となるデータを準備**

SAS Contextual Analysisの用途は、非構造化データから価値を引き出すことだけではありません。最新の対話操作型のアナリティクス環境に取り込んで探索や分析を行うためのデータを準備することもできます。こうしたデータは、例えばビジュアル・アナリティクス環境の導入や、他のアプリケーション向けのメタデータ管理のための基礎として利用できます。

製品概要

非構造化データの扱いは簡単ではありませんが、適切な技術があれば大きな価値を引き出す可能性があります。現在のように変化のペースが速いビジネス環境では、非構造化情報の調査を怠ると、顧客の懸念、市場の動向、潜在的な問題の前兆といったビジネスに関する重要な洞察を見落とす恐れがあります。競争優位性を維持するには、この種の情報の取得が不可欠です。

しかし、無限に増え続ける文書の山を詳しく調査するための時間とリソースを確保できる企業や組織は多くありません。また、たとえ調査できたとしても、人為的なミス避けることはできません。今日のように豊富なデータを扱う業務環境に求められているのは、処理パワーが優れているのみならず、直感的に利用でき、機能性が高く、柔軟にカスタマイズできるようなソリューションです。SAS Contextual Analysisは、まさにこの要件を満たすテキスト分析ソフトウェア製品であり、競争優位性につながる重要な洞察をもたらす総合的な環境を提供します。

ガイダンス付きのテキスト分析モデルの開発／展開を1つのシステムで一元管理

一元管理型の環境で言語学的なテキスト分析モデルの開発、改良、展開のすべてを行えるため、さまざまな作業工程の管理や、文書の分類方法の管理が簡素化します。また、統一された使いやすしいインターフェイスで全てを操作できるため、より柔軟かつ的確にシステムのパラメータをコントロールできます。

ハイブリッド型の文書分類アプローチ

機械学習は構文解析によって共通テーマ、カテゴリ・ルール、ドキュメント・センチメントを発見しますが、そこに各分野の専門知識を組み合わせることで分析をカスタマイズすることができます。このハイブリッド方式なら、より多くの文脈的な意味を発見でき、精度の高い結果が短時間で得られます。

自然言語処理

手作業で文書の内容を確認するのは、非常に時間がかかるのみならず、ミスも生じがちです。データ主導型の自然言語処理を導入すると、単調な手作業による定義が不要になります。このソリューションでは、ユーザーが入力データソースを指定すると、自動的に用語（単語または熟語）の出現頻度、品詞、同義語、語幹が検出されます。

初期タクソノミー（分類法）を開発するためのトピック検出機能

手作業で多数の文書から中心テーマを特定することは不可能ではありませんが、相当な困難が伴います。SAS Contextual Analysisは、これを自動的に行う機能を備えており、機械学習と統計分析の手法を用いて文書コレクションの中心テーマを特定します。

設定調整可能なルールの生成により、言語学的な文脈判定を改善

SAS Contextual Analysisでは、ブール演算子ベースのルール定義およびタクソノミーを開発する作業も簡素化されており、特許出願中の手法を用いて自動的にカテゴリ・ルールを作成することができます。そして、土台となるルール群を確立した後は、コンセプトの定義を通して、各ルールを意味論的に強化および改良することができます。

SAS®との直接的な連携

テキスト分析を通してビジネスに関する重要な洞察を発見することは、それだけで価値のあることですが、さらに深い分析を行うための第一歩でもあります。SAS Contextual Analysisから得られるテキストベースの豊富な結果は、SAS Visual Analyticsなどを用いてさらなる探索／分析を行うための基礎データとしても役立ちます。

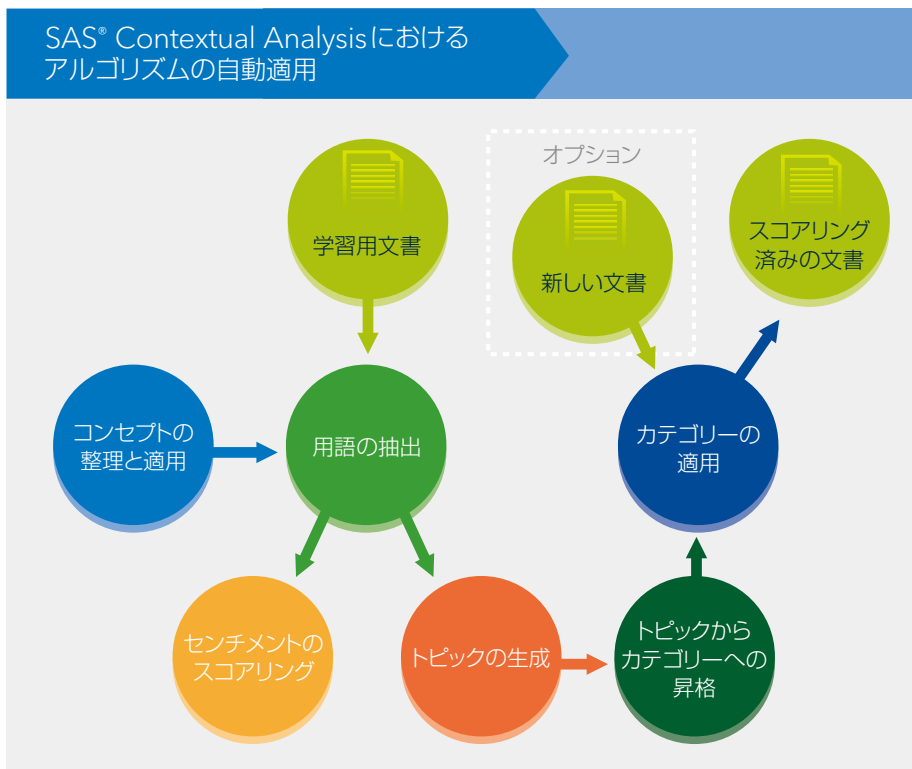


図1. 機械学習に基づいて生成された結果を分析担当者がさらに改良できる

主な機能

テキスト分析モデルのガイダンス付き開発・展開環境を提供する一元管理システム

- ポイント&クリック方式で操作できるGUI環境による自然言語処理。以下の処理に対応：
 - 用語（単語または熟語）
 - トピックの発見
 - カテゴリーの定義
 - コンセプト（概念）の規定
 - 文書レベルのセンチメントの特定
- テキスト分析モデルの調査から改良、導入までをカバーする、ガイド付きのテキスト分析モデル開発環境
- プロジェクトの開始点として、SAS Content Categorization から既存のタクソノミー（分類法）をインポートすることが可能
- カスタム・コンセプトの作成／変更／有効化（無効化）と、言語学的ルール定義のテスト（妥当性チェックを伴う）を同一の対話操作型インターフェイス内で実行
- 初期のプロジェクト定義のために利用できる、構築済みのプロジェクト・ウィザード
- テキスト分析モデルの開発工程全体で分析担当者をガイドするインライン・ヘルプ
- テキスト分析モデルの作成工程の各段階のステータス、処理状況、メッセージを表示するプロパティ・パネル。モデルの開発中に生じた問題の診断を支援（例：トピックの生成に必要な数の文書が文書コレクションに存在しない）
- プロジェクト管理のための各種オプション。複数のテキスト分析モデルを同時に開発・運用したり、各モデルの開発ステータスを示す説明を表示したりすることが可能
- プロジェクトの全てのプロパティについてメタデータを一元的に管理

ハイブリッド型の文書分類アプローチ

- 自動化された自然言語処理。用語の発見（スタート・ワードリスト／ストップ・ワードリストでカスタマイズ可能）、同義語の検出、スペルチェック、ステミング（語幹抽出）などが可能
- ドラッグ&ドロップ操作によるカスタマイズ。保持または削除する親用語をウィンドウ間で簡単に切り替え可能
- SAS Text Minerによるトピック発見を、機械学習で生成されるタクソノミー（分類法）の初期化に利用可能
- データソースを後で使用するオプション。タクソノミーを最初に構築し、文書コーパスに適用することが可能
- トピックに含まれる用語を視覚的な操作で改変し、類似のトピックをマージして生成結果を改良することが可能
- トピックをカテゴリーに昇格させる際に、データ主導型のカテゴリー・ルールの初期定義のためのルールを生成
- あらかじめ構築済みの多数の言語学的演算子を用いてシステム由来のカテゴリーを編集し、ルールをさらに改良
- 一般的なエンティティ（地名など）について事前定義済みの17個の設定調整可能なコンセプトを用いてカスタム・コンセプトを構築し、文脈感知の精度を強化
- 特定の評価のためのコンセプトを有効化／無効化することが可能
- 定義済みのコンセプト・ルールの一部または全部について、ルールの構文を検証
- トピック発見過程で生成された文書（または特定のカテゴリーに関して選定された文書）について、文書レベルのセンチメントをデータ主導方式で自動的に特定

自然言語処理

- このソフトウェアの自然言語処理機能における自動化の対象：
 - 解析
 - トークン化
 - 形態素解析（および品詞タグ付け）
 - 同義語検出
 - スペルチェック
 - ステミング（語幹抽出）
- 削除／保持する用語のリストと用語の出現頻度を出力
- 「保持」および「削除」ウィンドウ間でドラッグ&ドロップ操作が可能（親用語のみ）
- 追加のテキスト処理として、スタート・ワードリスト／ストップ・ワードリストを適用（して用語を処理の対象に包含、または対象から除外）することが可能。この製品には、通常は分析に適さないとされる一般的な用語（冠詞や接続詞）に関する仮のストップ・ワードリストが含まれています。分析担当者は、この事前に定義されたストップ・ワードリスト（現時点では英語のみ）を編集できます。
- ユーザー定義のスタート・ワードリスト／ストップ・ワードリストの使用も可能

主な機能(前ページより続く)

自然言語処理(続き)

- 入力の SAS データ内の対応する列を定義することにより、カスタム定義のカテゴリーを含めることも可能
- 専用パネルでカスタム・コンセプトを追加することも可能(定義済みの演算子構文を用いて記述)
- テキスト分析担当者は、入力文書を使わずにタクソミー(分類法)だけを先に開発し、後からコーパスに対してタクソミーを実行することも可能

初期タクソミー(分類法)を開発するためのトピック自動検出

- 機械学習による自動検出により、関連付けられた的確性スコアに基づいて入力文書コレクションから中心テーマを特定
- トピック内の用語の関連性をワードクラウド(閾値の設定調整も可能)や対話操作型の用語マップで調査。また、トピックの詳細をドリルダウンして関連性を評価し、トピック発見の精度を改良
- 文書とそのテーマとの関連が分かるような形で、トピックレベルのセンチメントを視覚的に表示
- 分析担当者は、トピックを2つに分割(必要に応じて分割を繰り返すことも可能)、あるいは類似のトピックを1つに併合することにより、生成されるトピックの数をコントロールすることが可能
- そうすることが望ましいトピックをカテゴリーに昇格させて分類ルールを定義した上で、コンセプトの定義によってトピックを補強

設定調整可能なカテゴリー・ルールの生成

- ユーザー定義のトピックに基づき、カテゴリー・ルールの初期定期を自動的に生成
- 理解しやすいプール演算子ベースの定義により、カテゴリー・モデル(つまり、タクソミー)を作成
- システム生成ルールの堅牢性が、真陽性、偽陽性、偽陰性という診断の形で提示
- カテゴリー・センチメントの視覚的な表現を文書ビュー内に生成
- カスタム・カテゴリーと同様、ルールもゼロから編集/補強/削除/定義することが可能
- 分析担当者が詳細なルール・モデルの仕様を記述するために使用できる構築済みのルール演算子を数多く用意
- 分析担当者は、構築済みのカテゴリー階層(つまり、タクソミー)を、入力データセットの定義済みフィールドから、または既存の SAS Content Categorization Studio プロジェクトからインポートすることが可能
- ルールの展開に先立って、入力データセットを用いてテストすることが可能

コンセプトによる文脈判定の改善

- コンセプト(概念)によってカテゴリーに文脈的な特異性を付与してモデルを改良し、より正確な意味を抽出
- 定義済みのコンセプトを利用できるため、ルールの記述が不要。日付、場所、時間などの一般的なエンティティを定義するコンセプトを標準で用意
- 既存の SAS Content Categorization Studio プロジェクトから、既存のコンセプトをインポートすることも可能
- カスタム・コンセプトは、定義済み演算子のセットを用いて記述することが可能
- 定義済みのコンセプト・ルール、既存のコンセプト・ルールのどちらも必要に応じて編集・改良することが可能
- コンセプト・ルールの検証機能。ボタンをクリックするだけで、構文の問題点を詳しく説明した診断エラー・メッセージを表示

SAS® との直接的な連携

- SAS Metadata Server との連携(SAS データセットをテキスト分析に直接入力する機能を新たに搭載)
- SAS Visual Analytics Hub との直接的な連携により、結果を対話型操作で探索することが可能
- 文脈依存モデルを他の任意の SAS アプリケーションに適用するモデルスコアリング・コードのダウンロード機能。これには、センチメントおよびコンセプト・スコアコードを表示してダウンロードする機能、および、他の文書のスコアリングに使うカテゴリー・スコアコードが含まれる

SAS Contextual Analysis の詳細については、sas.com/jp/go/contextual-analysis をご覧ください。



THE
POWER
TO KNOW.

SAS Institute Japan 株式会社 www.sas.com/jp

jpnsasinfo@sas.com

本社 〒106-6111 東京都港区六本木6-10-1 六本木ヒルズ森タワー 11F
大阪支店 〒530-0004 大阪市北区堂島浜1-4-16 アクア堂島西館12F

Tel: 03 6434 3000 Fax: 03 6434 3001
Tel: 06 6345 5700 Fax: 06 6345 5655

このカタログに記載された内容は、改良のため予告なく仕様・性能を変更する場合があります。あらかじめご了承ください。

SAS、SAS ロゴ、その他の SAS Institute Inc. の製品名・サービス名は、米国およびその他の国における SAS Institute Inc. の登録商標または商標です。その他記載のブランド名および製品名は、それぞれの会社の商標です。Copyright©2015, SAS Institute Inc. All rights reserved.

JP2015FS_CA_SE