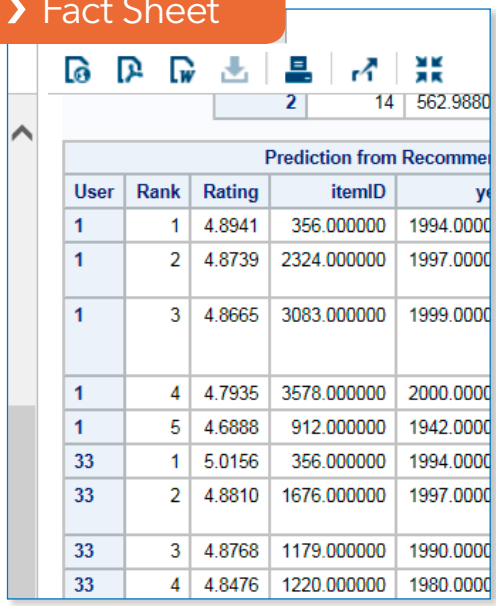
› Fact Sheet



		2	14	562.9880
			Prediction from Recommer	
User	Rank	Rating	itemID	y
1	1	4.8941	356.000000	1994.0000
1	2	4.8739	2324.000000	1997.0000
1	3	4.8665	3083.000000	1999.0000
1	4	4.7935	3578.000000	2000.0000
1	5	4.6888	912.000000	1942.0000
33	1	5.0156	356.000000	1994.0000
33	2	4.8810	1676.000000	1997.0000
33	3	4.8768	1179.000000	1990.0000
33	4	4.8476	1220.000000	1980.0000

主な機能

SAS® In-Memory Statistics for Hadoopは、データ・サイエンティストのためのインタラクティブなアナリティクス環境です。分析用データの準備から、変数の変換、探索的分析の実行、モデルの作成と比較、モデルのスコアリングまで、ビッグデータ・アナリティクスのすべてをHadoop環境内で行うことが可能です。

ビジネスメリット

高速かつ強力でカスタマイズ可能なインメモリ・プログラミング言語を提供します。この言語では、Hadoopに保管された大量のデータを対話操作型の環境で複数のユーザーが同時に分析できます。そのため、分析担当者の生産性が向上し、時間の余裕がない状況でも創造性を発揮することができます。

対象ユーザー

大規模で複雑なデータをHadoopで分析する必要のある統計担当者、データマイニング担当者、データ・サイエンティスト、エンジニア、研究担当者、バイオ統計担当者、科学者向けに設計されています。

SAS® In-Memory Statistics for Hadoop

Hadoop上でハイパフォーマンスな分析を行うための対話操作型アナリティクス環境

異種混在のデータストレージにHadoopを組み込む企業や機関が増えています。しかし残念ながら、Hadoopの登場後も、大量のデータから瞬時に洞察を得るための分析処理を超高速で実行するのが難しい、という状況が続いていました。また、分析ライフサイクルの各行程を適切に管理するためには複数のソフトウェア製品と希少なデータ・サイエンティストを揃える必要があることも、大きな課題でした。Hadoopから的確なタイミングで知見・洞察を導き出すためには、これまでとは異なるアプローチが必要です。

SAS In-Memory Statistics for Hadoopは、複数ユーザーが同時に膨大なデータを分析できる環境を提供します。この環境では、非常に強力な分析手法との組み合わせにより、高度な分析にもとづく意思決定のためにHadoopデータを活用し、価値の高い洞察を極めて高速に導き出すという画期的な手法が実現します。モデルの開発時間も大幅に短縮されるため、より多くのモデルをより迅速に現場に展開することが可能です。

また、このソリューションでは処理時間が最小限に抑えられるため、全社規模のニーズにも余裕を持って対応できます。つまり、より多くのユーザーがより多くのデータを扱い、より複雑な課題の解決に取り組むことが可能になるのです。

利点

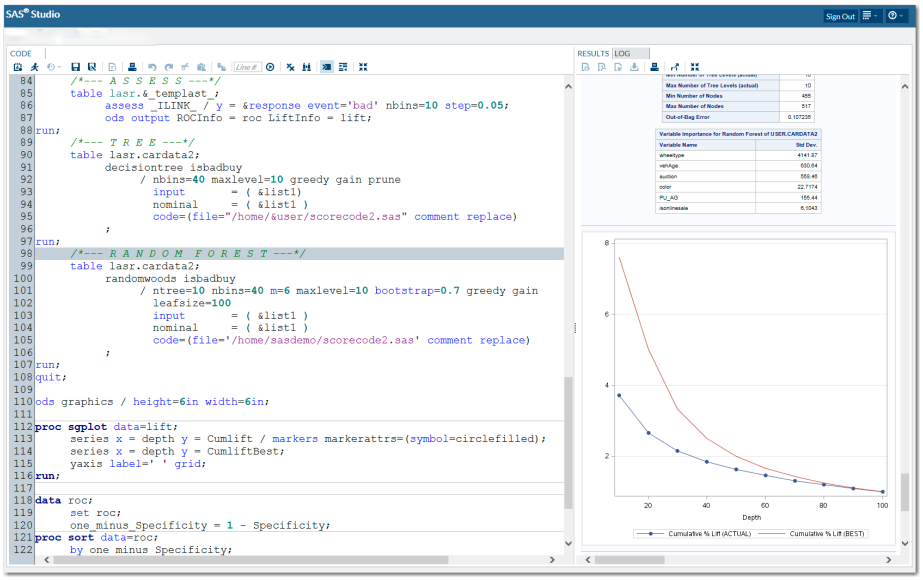
- **Hadoopを深く探索し、正確な洞察を高速に獲得**
  最先端の統計アルゴリズムや機械学習技術を駆使して最良の答えを見つけることができます。複数の分析アプローチを探索・活用して洞察を導き出し、事実にもとづく意思決定を行えるようになります。
- **データ・サイエンティストの生産性を向上**
  高速なインメモリ・アナリティクス・プログラミング言語を用いて、複数ユーザーが同時に、かつインタラクティブにHadoop内のビッグデータを分析することが可能です。データの準備、操作、変換、探索、モデル作成、アクセス、スコアリングのすべてをHadoop内で実行できます。
- **拡張性の高い環境を最大限に活用**
  これまでは、統計担当者やデータ・サイエンティストがHadoop環境でデータを準備、モデリング、スコアリングするには、異なるプログラミング言語や製品を組み合わせる必要がありました。また、モデル運用の段階では、ソフトウェアの規模面の拡張性に難点がありました。しかし、これからは違います。SASのソリューションは、データの操作や探索からモデルの構築と展開までを通じて豊富な実績があり、効果検証済みの正確な結果を提供します。さらに、お客様の本稼働環境に合わせて運用の規模を拡張することができます。
- **データ処理における不要な行程を回避**
  Hadoop上で動作するSASのインメモリ・インフラストラクチャでは、コストのかさむデータ移動が不要となり、データは分析セッションの終了時までメモリ内に保持されます。これにより、データ処理の待ち時間が大幅に削減され、超高速での分析が実行可能となります。

sas
THE POWER TO KNOW®

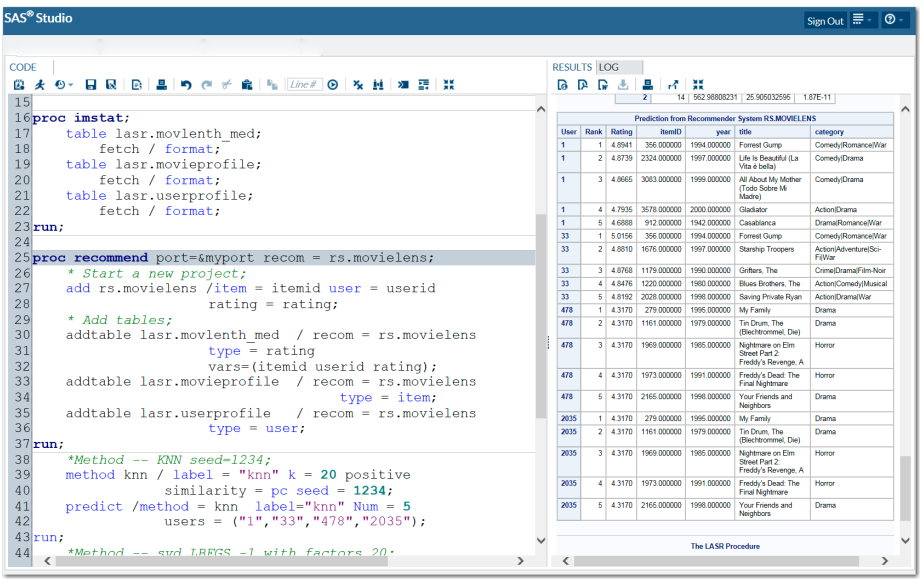
概要

SAS In-Memory Statistics for Hadoopは、Hadoopを活用したアナリティクス・ライフサイクル全体をカバーする、インタラクティブなアナリティクス環境です。データの管理、変数の変換、探索的分析の実行、モデルやスコアの構築／比較といった処理を、複数のユーザーが同時に実行することができます。インメモリ・アナリティクス処理を採用しているため、ほぼ瞬時に結果が得られます。

回帰、クラスタリング、決定木やランダムフォレストといった最先端の統計手法や機械学習技術を用いて、Hadoop上に保管された複雑なビッグデータに潜む新たなパターン、傾向、関係を特定することができます。インタラクティブ言語と標準装備のアクションを利用することで、MapReduceやRなどの他のプログラミング言語と比べて短時間で複数のモデリング・アプローチを探索し、最適解を明らかにすることができます。その結果、より優れた戦略のもとで業務プロセスの改善に取り組めるようになります。



ランダムフォレスト手法の使用例。高度なインタラクティブ性を備えたプログラミング環境で、非常に幅広い統計アルゴリズムや機械学習技術を利用できます。



自由度の高いカスタマイズを通じて、パーソナライズされた有用なレコメンデーションをリアルタイムで作成できます。

インタラクティブ・プログラミング環境

同時に複数のユーザーがインタラクティブな操作でHadoopの大量データを分析することができ、さらに、大量データを用いた複雑な分析でも、瞬時に答えを得ることができます。高度かつ使いやすいインタラクティブ環境であるため、思考の流れに沿ったアドホックで柔軟な作業が可能となり、ビジネス上の付加価値の高い複雑なWhat-If分析も可能となります。

インメモリ・アナリティクス処理

従来課題であった、分散Hadoopクラスター環境におけるマルチパスアクセスを解決し、完全にスケーラブルな分散並列型インメモリ・テクノロジーを実現しています。これにより、Hadoop上で高速なアナリティクス処理（特に機械学習処理）が実行可能です。このため、予測モデルの開発時間を大幅に短縮でき、より多くのモデルをより迅速に展開し、現場で活用できるようになります。

データを一貫してインメモリで処理（一度のディスクアクセスで何度も分析処理を実行可能）

このソリューションでは、データを並列的に1度ロードするだけで、複数ユーザーによるインタラクティブなデータ探索や高度なアナリティクスのすべてをサポートすることができます。全データが常にメモリ内に保持されるため、処理が超高速で実行され、待ち時間が大きく減ります。

アナリティクスのためのデータ管理

予測モデリングのためのデータ準備である、データ統合、変数加工や新規変数の作成、探索的分析の実行といった操作もインメモリかつインタラクティブな環境で実行可能です。Hadoop上でのデータ準備作業が大幅に効率化され、予測モデルの開発と業務への適用をより少ない時間で可能にします。

ハイパフォーマンスな統計解析と機械学習技術

高度かつ多様なアナリティクス・アルゴリズムを駆使し、従来よりも大幅に短い時間でパターンや傾向、異常、重要な変数や関係性を探り出すことができ、よりよい意思決定をより迅速に行うことが可能になります。業界随一の網羅的な統計解析アルゴリズムおよび機械学習技術を取り揃え、それらを単一の操作性で提供するのはSASだけです。

テキスト・アナリティクス

非構造化（および構造化）データを様々なテキスト解析技術を用いて分析することが可能です。構造化データとテキストデータを組み合わせて使用することで、これまで見つけることができなかった関係性を明らかにし、予測モデルの精度をさらに向上させます。

効率的な予測モデルの開発作業

予測モデルの作成、比較、評価を高速に行うことが可能です。複数のアナリティクス手法を用いて、思考のスピードに沿ったモデル開発作業を進めることができ、より効果的で迅速な意思決定が可能となります。例えば、モデルのコンポーネントを変更した場合の影響をリアルタイムで確認できるほか、計算項目の追加・削除、フィルター条件の変更、変数の入れ替え、連続値からカテゴリー値への変換といった、モデル開発に関わる試行錯誤を迅速に行うことができます。

レコメンデーション・システム

非常に自由度の高いカスタマイズを通じて、パーソナライズされたリアルタイム・レコメンデーションモデルを作成できます。協調フィルタリング、行列の分解、ハイブリッドモデル、マーケットバスケット分析（アフィニティ分析）を用いて、「次に提示すべき最良のオファー」を明らかにすることができます。このような独自のレコメンデーション・システムを構築することで、売上の最大化を図ることができます。

主な機能

インタラクティブなインメモリ・プログラミング

- すべての数学的計算をインメモリで実行
- ソートやインデックスの利用を不要とするダイナミックなグループ処理
- 新たなWebベースのプログラミング環境 SAS® Studio
- インタラクティブなプログラミング言語環境により、処理の実行、結果の取得、そしてさらなる処理のアドホックな実行が可能
- データの再読み込みやディスクへの一時書き出しが不要で、すべてのアナリティクス・タスクを一つのインメモリデータに対して実行可能
- 新規項目の作成や加工、フィルタリング、グループ処理が可能

アナリティクスのためのデータマネージメント

- HDFS（Hadoop Distributed File System）の活用
- より効果的なデータアクセスのための柔軟なパーティショニング機能
- 一時テーブルを作成し他のユーザーに公開可能
- フィルターや結合、変数の加工などのデータ加工機能
- Update／Append／Set が可能。フィルターや変数の加工、集計
- クライアント側での二次加工のための結果セットテーブルのエクスポート機能
- 全ての数学的計算をインメモリで実行
- ソートやインデックスの利用を不要とするダイナミックなグループ処理

統計アルゴリズムと機械学習技術

- ディシジョンツリー
- 時系列予測（フォーキャスティング）
- 一般線形モデル
- 一般化線形モデル
- ロジスティック回帰
- ランダムディシジョンフォレスト
- クラスタリング（k平均法）
- クラスタリング（DBSCAN）
- アソシエーション

記述統計

- Distinct Count
- 箱ひげ図
- ピアソン相関
- クロス集計（重みづけ含む）
- 関連度の計算を含む、各種の分割表
- パラレルグループ処理
- ヒストグラム（ビン化、最大値の閾値、外れ値などの機能付き）
- 1回のデータパスで多次元サマリーを作成
- 複数の変数に対して百分位を計算
- 各種の要約統計量。オブザベーション（観測対象）の数、欠損値の数、非欠損値の合計、平均、標準偏差、標準誤差、修正／無修正平方和、最小と最大、変動係数など
- 正規／tricube／二次カーネル関数を用いたカーネル密度推定

SAS In-Memory Statistics for Hadoop の詳細、ホワイトペーパーのダウンロード、スクリーンショットの確認、関連資料の閲覧については、Web サイトをご覧ください。
sas.com/jp/go/imstat

主な機能（続き）

予測モデルのアセスメント
- リフトチャート、ROC チャート、誤分類表、一致統計量などを使用した一般的なモデル比較手法

テキスト解析
- パースとステミング
- スタート・ストップ ワードリスト
- 単語と文書の頻度分析
- 特異値分解（SVD）
- エンティティの抽出と分解
- 文書が意味するトピックの推定

予測モデルのスコアリング
- SAS DATA ステップコードの生成
- 学習データ／ホールドアウトデータ／新規データへのスコアリング・ロジックを適用するためのスコアコード

レコメンデーション・システム
- インタラクティブな RECOMMEND プロシジャ。すべてのアルゴリズムをインメモリで実行可能
- フィルターを対話操作で適用して特定の対象者に向けたレコメンデーションを開発
- ユーザー、アイテム、レーティングテーブル（評価表）のメモリへのロードをプロジェクトベースでサポート
- 履歴を使わない加重平均に基づく新規ユーザーを対象としたコールドスタート
- シンプルなベンチマークとして使用されるスロープワン（Slope-One）高速回帰ベースのモデル
- k 近傍法（類似度として余弦距離、修正余弦距離およびピアソン相関を含む）
- 行列分解（損失関数、正則化定数、最適化手法などのオプションを利用可）
- 用語頻度やドキュメント頻度重みを含む他の属性を用いたユーザーおよび／またはアイテムのクラスタリング
- ハイブリッドまたはアンサンブル・モデル
- 学習や検証データの評価に対するユーザーおよびレーティング（格付け）のホールドアウト・セットの定義
- 新規ユーザーやテーブルのスコアリングを行った結果であるアクションを予測

PROC Recommend Movie Recommendations

Obs	UserID	Rank	Rating	Title	Year	Category
1	Mary Clarke	1	6.07	Indiana Jones and the Last Crusade	1989	Action|Adventure
2	Mary Clarke	2	5.54	Star Wars: Episode V - The Empire Strikes Back	1980	Action|Adventure|Drama|Sci-Fi|War
3	Mary Clarke	3	5.43	Alien	1979	Action|Horror|Sci-Fi|Thriller
4	Mary Clarke	4	5.40	Pretty Woman	1990	Comedy|Romance
5	Mary Clarke	5	5.30	Stand by Me	1986	Adventure|Comedy|Drama
6	John Thompson	1	4.70	Fugitive, The	1993	Action|Thriller
7	John Thompson	2	4.69	Elizabeth	1998	Drama
8	John Thompson	3	4.67	Saving Private Ryan	1998	Action|Drama|War
9	John Thompson	4	4.64	Philadelphia Story, The	1940	Comedy|Romance
10	John Thompson	5	4.64	Strictly Ballroom	1992	Comedy|Romance

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 Tel: 03 6434 3000 Fax: 03 3434 3001
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: 06 6345 5700 Fax: 06 6345 5655

JP2014FS_IMSfH_SE