

Big Data Module

ビッグデータ攻略法は
ここに!

本格的な分析を大規模データで実現

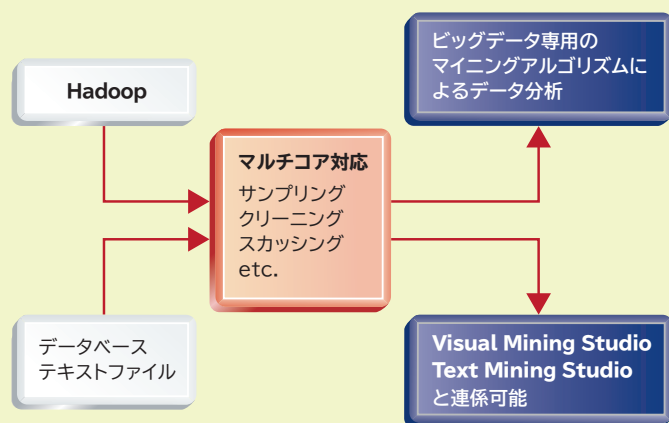
WEBアクセスログ、プローブデータ、取引データなど、テラバイト～ペタバイト級のデータセットが続々と出現しています。Big Data Moduleは、大規模データ分析のためにマイニングのスペシャリストである数理システムが自社開発した**ビッグデータ分析ツール**です。

集計だけでは満足できない

確率的勾配降下法を使用したオンラインマイニングアルゴリズムなど、高速かつ省メモリで動作する最新のアルゴリズムを多数採用しました。ビッグデータでも、単純集計だけでは終わらせず、高度な分析を実行できます。

直感的な分析インターフェース

読み込み、加工処理、分析などの処理はアイコン化されており、直感的なマウス操作だけで高度な分析を実行できます。



特殊な分析専用マシンは不要

冷蔵庫のような巨大で特殊な分析専用マシンを用意する必要はありません。市販の高性能サーバーマシンを1台用意していただければ、マシンパワーをフルに活用してテラバイトオーダーのデータのマイニングが可能です。

VMStudio、TMStudio、Hadoopとの連携

データマイニングシステム Visual Mining Studioや、テキストマイニングシステムのText Mining Studioなど、数理システム製品とのシームレスな連携が可能です。Big Data Moduleの分析結果をさらに掘り下げて、多様な分析が行えます。Hadoopとの連携により、ペタバイトオーダーのデータの分析が可能となります。

Python利用がより効率的に、便利に **NEW**

機械学習、Deep Learningで注目を浴びているプログラム言語Python。各種のライブラリが充実しているので、強力なデータ分析環境になりますが、一方でデータをインポート、加工、グラフ表示するにもプログラミングが必要です。

Visual Analytics PlatformやBig Data Moduleのデータインポート、加工機能を活用し、グラフ機能でデータを可視化、必要な分析は実績のあるPythonでプログラミング、そんなときにBig Data ModuleのPython連携機能が役に立ちます。



適用データ例

通信ログデータ、POSデータ、通信トラフィック、気象データ、プローブデータ、レセプトデータなど。
使用可能なデータ形式や上限サイズについてはお気軽にお問い合わせください。

機能詳細

並列処理

インポート、クリーニング、サンプリング、集計などを簡単に並列実行できます。CPUの処理性能を最大限に生かして分析を行うことができます。

オンラインマイニングアルゴリズム

オンラインアルゴリズムでは逐次的にデータの処理を行い、モデルの更新を行います。必要となるメモリはデータサイズに依存せずに非常に小さく、処理時間もデータサイズの線形オーダーで抑えられます。そのため、データのサイズを気にせず素早く分析を行うことができます。

機能一覧

■前処理

データインポート	csvファイルなどのテキスト形式のデータを取り込む
クリーニング	データの欠損値、異常値を除外する
サンプリング	ランダムサンプリングを行う
スカッシング	BIRCHアルゴリズムでデータの圧縮を行う
コアセット	k-meansに最適化したデータの圧縮を行う
データハンドリング	対話的にフィルタリング、集計などのデータ加工を行う

■分析

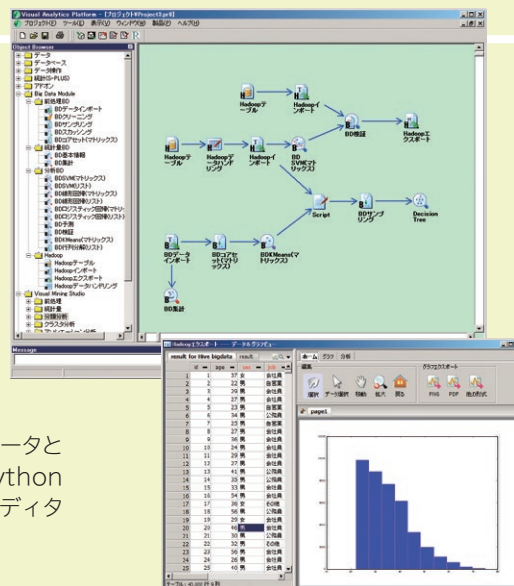
オンラインSVM	Multiclass Support Vector Machineで分類モデルを作成する
オンライン線形回帰	線形回帰で予測モデルを作成する Lasso、Elastic Netによる変数選択も可能
オンラインロジスティック回帰	多項ロジスティック回帰で分類モデルを作成する
予測	作成したモデルを使用して予測を行う
検証	予測値と実際の結果を比較してモデルの評価を行う
オンラインk-means	k-meansでクラスタリングを行う
オンライン行列分解	行列分解、非負値行列分解でクラスタリング、特徴抽出を行う
レコメンデーション	オンライン行列分解の結果を使用してレコメンデーションを行う

Hadoop関係

Hadoop上のデータの閲覧、データ加工、データの入出力をGUIから簡単に行うことが可能です。データがHadoop上にあることを気にせず分析を行うことができます。

Python関係 NEW

Visual Analytics Platformでデータインポート、加工したデータをPythonで利用可能なデータとするデータ変換アイコン、またPythonプログラミングをサポートするエディタ機能を搭載しています。



■統計量

基本情報	行数、列数、最大値、最小値などの基本的な情報を表示する
集計	平均、分散、最大値、最小値などの統計量を計算する 重み付き、キー付きの集計にも対応

■その他

データ&グラフビュー	データの閲覧、グラフの作成を行う
スクリプト	独自のデータ処理、分析内容を記述し実行する
バッチ処理	あらかじめ作成した分析フローをコマンドプロンプトから実行する

■外部関係

Python関係	Pythonスクリプトを実行する
Hadoop関係	Hadoop上のデータの閲覧、加工、集計、取り込みを行う
Visual Mining Studio関係	Big Data Moduleでデータ圧縮などを行った結果に対して、Visual Mining Studioの多様な分析アルゴリズムを適用する
Text Mining Studio関係	分かち書きや辞書処理といったテキストデータ特有の処理を行って文章をデータ化し、分析を行う
その他数理システム製品との関係	同一のGUI上でシームレスに関係が可能

動作環境

対応OS	Microsoft® Windows® 8.1 / 10 / Server 2012 / 2012 R2 / 2016 / 2019 いずれも64 bit版のみ
CPU	2.0 GHz以上 コア数2以上 (2.8 GHz以上、コア数4以上推奨)
メモリ	4 GB以上 (2 GB×コア数以上推奨)
HDD空き容量	10 GB以上 (推奨空き容量は分析対象データサイズに依存)

推奨環境は目安です。データサイズ、分析内容によって処理時間とハードウェア要件は大きく異なりますので、お気軽にご相談ください。

※Microsoft、Windowsロゴは、米国Microsoft Corporationの米国およびその他の国における登録商標または商標です。
※会社名、商品名は各社の商標または登録商標です。

NTT data 株式会社NTTデータ 数理システム

〒160-0016 東京都新宿区信濃町35番地 信濃町煉瓦館1階 TEL 03-3358-6681 FAX 03-3358-1727

(e-mail) bigdata-info@msi.co.jp (URL) https://www.msi.co.jp/bdmodule/