



[NTT HOME](#) > [NTT持株会社ニュースリリース](#) > [2011年](#) > ビッグデータのリアルタイム処理を可能にする大規模データ分析基盤技術を世界に先駆けて開発

NTT持株会社ニュースリリース

(報道発表資料)

2011年10月26日

日本電信電話株式会社

株式会社プリファードインフラストラクチャー

ビッグデータのリアルタイム処理を可能にする大規模データ分析基盤技術を世界に先駆けて開発 ～10月27日よりオープンソースソフトウェアとして公開～

日本電信電話株式会社(東京都千代田区、代表取締役社長:三浦 愷、以下「NTT」と)と株式会社プリファードインフラストラクチャー(東京都文京区、代表取締役社長:西川 徹、以下「PFI社」)は、ビッグデータ^{*1}と呼ばれる大規模データをリアルタイムに高速分析処理する基盤技術「Jubatus^{*2}」(第1版)を開発しました。

直近の潮流変化を調べる等、対象データの新鮮さを意識したビッグデータ分析アプリケーションを作成するニーズに対して、従来のバッチ処理による定期的な一括処理は、新しいデータの分析が次の実行タイミングまで待たされる点で十分ではありませんでした。

Jubatusは、新しいデータを取り込みリアルタイムに分析することで、不正検知、市場・経済・株価予測、自然災害予測、製造業の部品・材料調達予測、健康リスク予測、生命科学・自然科学上の予測技術などの広い領域で付加価値サービスの創出に貢献します。

今回の開発は、NTTの情報流通プラットフォーム研究所とPFI社とのオープンイノベーションの成果であり、10月27日よりJubatus OSS 提供サイト <http://jubat.us/>にて公開し、ビッグデータ活用に寄与する共通財産として、オープンソースソフトウェアとして公開します。

技術の背景

近年、さまざまな分野で「情報爆発」という言葉で語られるデータの爆発的な増加が顕在化し、ビジネスにおけるビッグデータの積極的な有効活用がますます重要となってきています。

現在のビッグデータの解析方法では、クラウド環境を構成するサーバ群にビッグデータをいったん蓄積し、一定期間分をまとめて高速に処理するバッチ処理が一般的です。特にHadoop^{*3}の注目度が高く、利用実績も増えています。

ところが、世の中の変化のスピードは極めて速く、SNS分析、異常トラフィック検知や不正アクセス検知などのアプリケーション領域で高度な解析・予測に基づく迅速な意思決定を実現するために、時系列で流れ込んでくる大量データを、蓄積せずリアルタイムに、高度解析する技術が必要となってきました。

技術の概要

Jubatusは、「大量データ」を「常に素早く」「深く解析」することを狙った(図1^①)、大規模分散リアルタイム解析フレームワークです。Jubatusでは「大量データ」を「常に素早く」処理するために大量のデータを複数のサーバに振り分け並列かつ逐次的に処理させていきます。また、「深い解析」には、高度な統計処理や機械学習が必要となり、それを分散環境で実施するためには、複数サーバがお互いの途中処理結果を共有する仕組みが必要となります。この共有にはサーバ間通信によるボトルネックの発生リスクが伴うため、通信方法を工夫しないとトータルの性能が低下します。

そこで、Jubatusでは、複数のサーバ間で緩やかに途中処理結果を共有することにより、サーバ間の通信オーバーヘッドの削減や安定性の向上を実現し、高いリアルタイム性と解析精度を確保しています。

技術の特徴

Jubatusは、主に以下の特徴を有しています(図2□)。

(1) MIX処理方式

次の3つの機能から構成される処理方式です。

- <1> MIX計算: データ解析処理のロジックに応じて、集約計算ロジックをアレンジする機能
- <2> MIXプロトコル制御: サーバ間で、途中の解析結果を答え合わせする際のデータの集め方と再分配のやり方を決める機能
- <3> メンバシップ管理機能: 絶え間なくデータを処理するため、サーバ障害をリカバリしたり、データが溢れるまえにサーバを追加したり、リーダを決める機能

同時並列で解析しても、全サーバが毎回待ちあわせて中間結果を集約することは明らかなボトルネックを生じます。毎回でなく、適度な時間間隔で他サーバと中間結果を交換して準標準化することで、各サーバはなるべくブレーキをかけずに自律的に走ることができます。アプリケーション要件が許容可能な範囲で集約結果の精度と厳密性(全体の一貫性)を緩めるように、リアルタイム性とスケーラビリティの両立を調整します(図3□)。

(2) プラガブル構成

共通インタフェース規定により、解析エンジン、解析モジュール、データストレージ方式(ローカル・分散)の柔軟な組み換え(プラグイン・アウト)を可能にしています。

(3) ワークフロー化

データ入力から、応用解析処理、解析エンジンなどのプロセス構成間のパスおよび並列実行などの柔軟で容易な定義と実行制御を可能にしています。

今回は、深い解析事例としてオンライン機械学習における多値分類器を実装し、これらの方式の確認と性能測定、チューニングを実施しました。

今後の展望

NTTとPFI社は、研究開発をさらに進めるとともに、Big Dataに関わる情報処理技術の発展に寄与するために、Jubatusコミュニティの拡大、ならびにその上のビジネスマーケットの拡大も目指し、リアルタイム型大規模データ分析基盤技術の普及促進とビジネス展開を目指します。具体的には、次のような応用サービスへの適用が考えられます。

「SNS分析アプリケーション」

日々大量に発生するSNSデータに対し、カテゴリ分け、曖昧検索、スパムフィルタリング、重要度判定などの高度な分析をリアルタイムに行うことにより、マーケティングなどに活用する。(図4□ Jubatusを適用したSNS分析アプリケーションのイメージ)

その他の用途として「センサーデータ解析」、「POSデータ解析」、「Logデータ解析」、「金融データ解析」、「行動解析」など適用分野は広範囲です。

参考

株式会社 プリファードインフラストラクチャー社(<http://www.preferred.jp>)について

PFI社は、自然言語処理技術、機械学習技術分野で優れた研究・開発技術者を有するベンチャー企業です。Jubatusにおける高性能な解析エンジンはPFI社の技術力から生まれました。

用語解説

*1 ビッグデータ

量が膨大であり、しかも構造が複雑であるため、従来の技術では管理や処理が困難とされているデータ群のこと。まだ明確な定義はないが、通常は、数百テラバイトからペタバイト級以上で、大量であると同時に、非定型であり、リアルタイム性の高いデータ群を指すことが多い。たとえば、RFIDタグや各種センサからの情報、ブログなどに代表される新しいコミュニケーションツールの記述なども含まれる。

*2 Jubatus

ネコ科の動物であるチーターの学術名「Acinonyx jubatus」に由来し命名した。オープンソースソフトウェアの公開により、「Jubatusコミュニティ」の形成を目指す。



「Jubatus」のシンボルマーク

*3 Hadoop

Google社の基盤システム (MapReduce、BigTable、GFS等) のオープンソースクローン。大規模分散処理基盤のバッチ処理系の代表的な存在であり、Hadoopコミュニティが成熟している。

別紙・参考資料

- ▶ [図1 Jubatusの位置づけ](#)
- ▶ [図2 アーキテクチャの概要](#)
- ▶ [図3 Jubatusの仕組み](#)
- ▶ [図4 SNS分析アプリケーションのイメージ](#)

本件に関するお問い合わせ先

■ NTT情報流通基盤総合研究所

企画部 広報担当

TEL : 0422-59-3663

E-mail : islg-koho@lab.ntt.co.jp

■ 株式会社 プリファードインフラストラクチャー

Jubatus担当

TEL : 03-6662-8675

E-mail : info@preferred.jp

ニュースリリースに記載している情報は、発表日時点のものです。現時点では、発表日時点での情報と異なる場合がありますので、あらかじめご了承くださいとともに、ご注意をお願いいたします。

[NTT持株会社ニュースリリース インデックスへ](#)

NTT持株会社
ニュースリリース

▶ [最新ニュースリリース](#)

▶ [バックナンバー](#)

▶ [English is Here](#)

NTT持株会社
ニュースリリース内検索

1997 ▼ 年 04 ▼
月 ~
2021 ▼ 年 11 ▼ 月

NTTグループの情報は
こちらからもご覧いただけます。



[▲このページの先頭へ](#)

[▶ 更新履歴](#) [▶ サイトマップ](#) [▶ お問い合わせ](#) [▶ 著作権](#) [▶ プライバシーポリシー](#) [▶ 情報セキュリティポリシー](#) [▶ ウェブアクセシビリティポリシー](#) [▶ 個人情報保護について](#)

Copyright © 2021 日本電信電話株式会社