

4



センサ・モバイルビッグデータを有効利用するためのプラットフォーム技術

原 隆浩（大阪大学）

センサやモバイルが生み出すビッグデータとその課題

✦ ICT 技術の 4 つのキーワード

総務省の平成 25 年度版・情報通信白書などでは、「モバイル」「ソーシャル」「ビッグデータ」「クラウド」が ICT 技術・サービス革新の要になると期待されている。この中で「ビッグデータ」は主に、M2M (Machine-to-Machine) や IoT (Internet of Things) と称されるように、センサなどの「モノ」がインターネットを介して相互に接続され、大量のデータが生成されることによって生じる新たな技術変革として注目されている。

特に、「モバイル」端末が生成する GPS や各種センサデータ、アプリやサービスの履歴データなどが、新たなビッグデータの主役と考えられている。また、「モノ」だけではなく、「人」が生成するデータ、特に「ソーシャル」ネットワークサービス (SNS) から生成されるユーザのメッセージや友人関係を表すグラフデータなどが貴重な解析対象として注目されている。さらに、ビッグデータ解析のためのプラットフォームとして、「クラウド」が注目されている。このように、先に述べたキーワードは、実は相互に密接に関係している。ビッグデータを中心に捉えると、ビッグデータを生成する新たな要因としてモバイル、ソーシャルが注目され、これを処理するためのプラットフォームとしてクラウドが注目されている。

✦ センサ、モバイルビッグデータの課題

上述のように、ビッグデータも、他の注目技術と相まって、さらに高度化する様相を成している。特に、本来のセンサから生成されるデータに加えて、モバイルやソーシャルから生成されるデータも、実社会を映し出すセンサデータ的一种として捉えることができる。このようなさまざまな情報源から生成されるデータを、それらが発生する機関内での利用にとどめず、広く共有するために、オープンデータ化の流れが進んでいることはよく知られている。

最近では、オープンデータ化からさらに発展して、多種多様なデータを統合的かつ効率的に処理・解析するためのプラットフォーム技術 (クラウドを含む) への要求が高まっている。

本稿では特に、センサやモバイルが生み出すビッグデータに注目し、それを有効利用するためのプラットフォーム技術について、最近の商用レベルのサービスおよび研究動向を紹介する。

商用プラットフォーム・サービス

国内外の多くの企業では、2012 年頃から、顧客情報やサービス利用履歴のいわゆる業務データ以外にも、センサデータなどを有効に解析することで、新たなサービスを創出できるものと考え、さまざまな製品開発が行われている。特に、ビッグデータの重要な特徴である Volume (大規模な量) と Velocity (高速レートの生成速度 (ストリーム))、Variety (多様性) が、一般の開発者がビッグデータ解析のためのシステムを構築する上で大きな障害

になることから、開発支援のためのツール群や、開発支援サービス・クラウドの提供が新たなビジネスとして盛んに行われている。

たとえば、東芝は統合ビッグデータプラットフォームとして、多種多様なセンサデータや企業内システム、ソーシャルメディアのデータをリアルタイムに監視しイベントを検知するためのイベント処理基盤 SmartEDA、ビッグデータを高速に蓄積・検索するためのデータストア基盤 GridStore、ビッグデータを高速に処理、解析するための並列分散処理基盤 GridData のミドルウェア群を提供している。特に、GridData では、一般的な並列分散処理基盤である Hadoop や NoSQL データベースの Hive、統計解析言語の R 言語と連携してビッグデータ処理を支援している。

NTT データは、Xrosscloud というコンセプトの下、M2M トータルソリューションとして、スマートフォン、センサ、自動車、業務データなどの連携・加工・分析による新たな価値の創出を支援することを目的としている。Xrosscloud の特徴は、プラットフォームだけではなく、アプリケーション自体の提供や、ビッグデータ分析・活用コンサルティングなど総合的な支援を行う点である。プラットフォームとしては、複数センサ処理やビッグデータ処理の基盤などがあり、アプリケーションとしては防災、施設監視、交通など典型的なものが提供されている。

INTEC では、ユビキタスプラットフォームとして、センサ、スマートフォン、オープンデータからのデータ収集・蓄積、および、ストリーム処理、ビッグデータ分析のための API 群を提供している。

これらの商用プラットフォームおよびサービスは、Hadoop や NoSQL などのオープンソースのツール群を利用して独自に開発する場合よりも、開発者の手間を削減できる。しかし、利用に費用がかかり、さらに、モバイルやセンサに起因するデータの特徴を考慮したデータ収集、処理、分析の機能の多くは自身で開発する（もしくは外注する）必要がある。そこで、これらを支援するための研究開発が盛んに進められている。

モバイル・センサデータ収集・共有のプラットフォーム

まず、モバイル・センサデータを収集・共有するためのプラットフォームに関する既存研究について紹介する。2003 年頃から、センサデータを共有するためのプラットフォームは盛んに研究されており、Intel リサーチらの IrisNet や、ブラウン大学らの Borealis、EPFL らの GSN (Global Sensor Network)、筆者らの X-Sensor などがある。

最近では、これらを発展させ、多種多様なセンサを統合利用するために、物理的なセンサを仮想センサとして定義して、クラウドサービスなどを展開する Sensor Cloud の研究開発が盛んに行われている。また、各種センサを搭載したスマートフォンなどの普及により、モバイルユーザが生成するセンサデータやメッセージなどを収集する参加型センシング（またはモバイルセンシング）が注目されている。本章では、これらのうちでプラットフォーム開発に主眼を置いた代表的な研究を紹介する。

✦ Sensor Cloud

Sensor Cloud と称するサービスやプラットフォームは複数存在するが、基本的にセンサ群が提供するサービスやデータを仮想センサと見なし、インターネットを介してクラウドサービスとして提供するものである。これに対して、ミズーリ科学技術大学 Sanjay Madria 教授らが開発する Sensor Cloud¹⁾ は、物理的なセンサデバイス上に仮想センサを構築し、Sensing as a Service を提供している。これにより、複数の機関が設置する異なるセンサネットワーク上のセンサ群を透過的に 1 つのセンサネットワークと見なして、データ収集や集約処理などの制御を行うことが可能となる。

この Sensor Cloud のシステムアーキテクチャを図-1 に示す。特徴としては、クライアント管理部、ミドルウェア部、センサ管理部の 3 つのレイヤから構成され、それぞれが独立していることである。センサ管理部は物理的なセンサの管理を行い、クライ

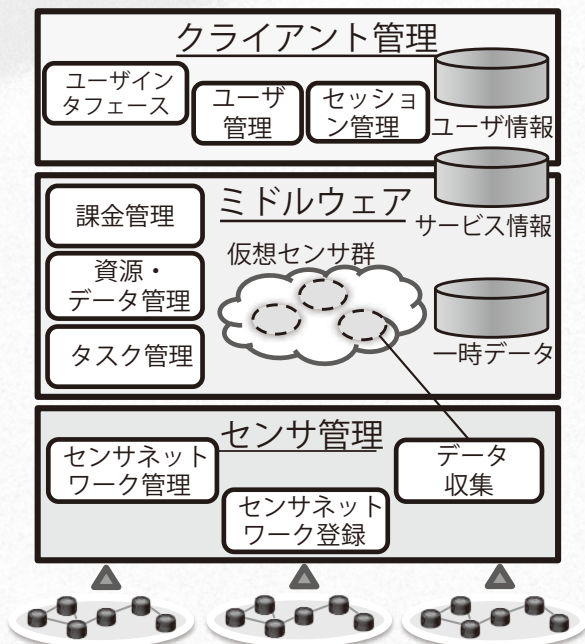


図-1 Sensor Cloud アーキテクチャ

クライアント管理部はクライアントや表示管理を行う。ミドルウェア部はデータ収集の制御などを行う。このように構成することで、アプリケーション開発者は、クライアント管理部とミドルウェア部のみの機能を利用して設計・実装を行い、センサネットワーク管理者はセンサ管理部のみの機能を利用することで自身のセンサ群をSensor Cloudに登録できる。

❖ Ohmage

モバイルユーザが生成するさまざまなデータをセンサデータとして収集する参加型センシングは、設置型のセンサでは実現できない以下のような役割が期待されている。

(1) 「人」に関する情報の解析（社会センシング）

ユーザ（人）自身がデータを生成するため、街中での人の動きのモデリングや、実社会のイベント、トレンドなどの発見が期待できる。

(2) 超高密度な環境センシング

都市部ではユーザが密に存在するため、数十[km]単位等で設置される気象センサと比較して、数十[m]単位等での詳細なセンシングが可能となる。

参加型センシングのプラットフォームに関する研究開発もいくつか進められており、カーネギーメ

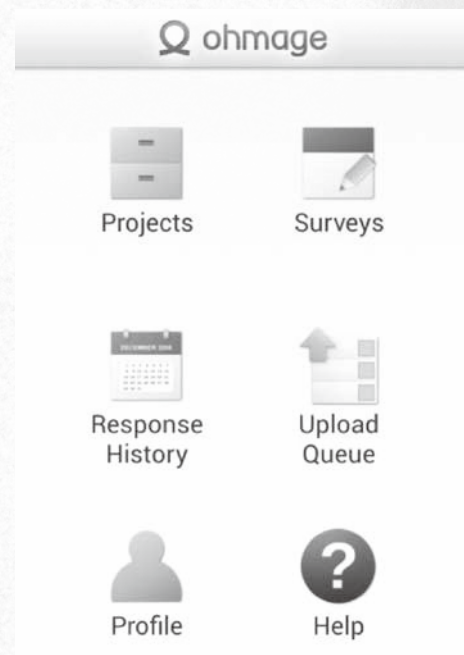


図-2 Ohmage アプリケーションのトップ画面

ロン大学らのSensr、ワシントン大学のODK Sensors、コーネル大学のOhmage、筆者らのSmasenseなどがある。本節では、これらの中で、Ohmageについて概説する。

Ohmage²⁾は、コーネル大学Deborah Estrin教授らが開発する参加型センシングのプラットフォームであり、Webベースのデータ管理や可視化、解析の役割を持つバックエンド（サーバ）と、ユーザ側のデータ収集用モバイルアプリケーションから構成される。図-2に、モバイルアプリケーションのトップ画面を示す。

Ohmageの顕著な特徴としては、クラウドソーシングのタスクを実行するように、ユーザがテキストやマルチメディアデータ形式で指示内容に対して回答するデータ収集方法（自己報告）と、モバイル端末のセンシングデータや端末利用状況（CPU・メモリ使用状況、通信履歴など）を自動で送信するデータ収集方法（受動データ）の両方を提供している点である。現状では、後者は、移動状況（静止・歩行・走行の種別と位置）、音声、システム利用状況の3種類を取得する機能のみが提供されているが、効率的に多様なデータを収集するために、他の参加型センシングのサービスとの連携についても検討が

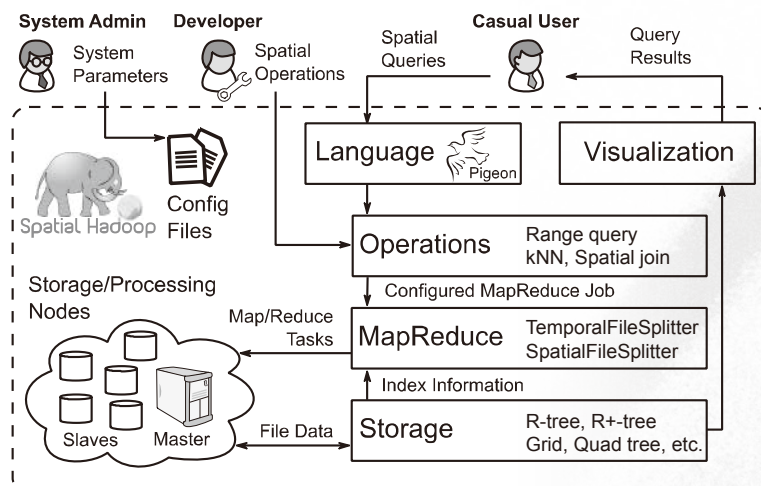


図-3 SpatialHadoop のシステムアーキテクチャ

行われている。

空間データの並列分散処理プラットフォーム：SpatialHadoop

モバイルユーザやセンサから生成されたデータは、一般に場所（や時間）に関連付けられることが多い。このようなデータは空間（Spatial）データと呼ばれ、環境モニタリング、人の行動解析、トレンド解析など、さまざまな応用で用いられている。そのため、大量の空間データを効率的に処理・解析可能なプラットフォームの重要性が増してきている。上記で紹介した商用システムを利用することも可能であるが、開発者が空間データを処理するための機能を独自に実装しなければならない場合が多く、負荷が大きい。そこで、空間データの並列処理プラットフォームに関する研究が盛んに行われており、エモリー大学の Hadoop-GIS や、NEC/UCSB の MD-HBase、ミネソタ大学の SpatialHadoop³⁾ などがある。これらの中で、総合的な機能の豊富さから、ミネソタ大学 Mohamed Mokbel 准教授らによる SpatialHadoop が特に注目されている。

✦ SpatialHadoop の特徴

SpatialHadoop は、Hadoop のシステムアーキテクチャにおける各レイヤに対して、空間データを処理するための機能を拡張する形で実装されてい

る（図-3）。SpatialHadoop の大きな特徴は、先述のように、提供する機能の豊富さである。さらに、Pigeon という独自の言語を採用しており、この言語は OGC（Open Geospatial Consortium）が定義しているデータ型や関数の標準に準拠している。一方、そのほかの既存のプラットフォームで採用されている言語で、OGC 標準に準拠しているものはほとんどない。

✦ SpatialHadoop の機能

本節では、SpatialHadoop のいくつかのレイヤについて、機能の概要を述べる。

■ Storage（ストレージ）レイヤ

Hadoop 自体はインデックス機能を有していないため、SpatialHadoop では、空間データの効率的な検索のために、Grid ファイルや R 木、R+ 木などの代表的なインデックス機能を提供している。

■ MapReduce レイヤ

Hadoop では、FileSplitter により入力ファイルを n 個の Split に分割するが、SpatialHadoop では上記のインデックスの情報を利用して、Spatial-FileSplitter と TemporalFileSplitter により入力ファイルを分割する。この際、時空間情報を考慮して、不要なファイルブロックの探索を枝符りし、処理時間の短縮を図っている。

■ Operations (操作) レイヤ

SpatialHadoop では、時空間データに対してよく用いられる範囲検索、k 最近傍検索、および空間結合処理の機能を提供している。

今後の展開

これまで述べたように、モバイルやセンサが生成するビッグデータを有効活用するためのプラットフォームに関する研究開発は盛んに行われており、今後もますます加速するものと考えられる。一方で、特にモバイルユーザが生成する位置情報や SNS データは、多くのプライバシー情報を含んでいる。今後は、プラットフォームにおいて、ユーザから提供されたデータのプライバシーを保護しつつ、提供するサービス・データの質を低下させない機能をさらに強化

する必要がある。さらに、複数のプラットフォームが所有するデータを横断的に解析しても、ユーザのプライバシーが侵害されないような仕組みなども検討する必要がある。

参考文献

- 1) Madria, S., et al. : Sensor Cloud : A Cloud of Virtual Sensors, IEEE Software, Vol.31, No.2, pp.70-77 (2014).
- 2) Tangmunarunkit, H., et al. : A General and Extensible End-to-End Participatory Sensing Platform, ACM Transactions on Intelligent Systems and, Vol.6, No.3, Article 38 (2015).
- 3) Eldawy, A. and Mokbel, M. F. : SpatialHadoop : A MapReduce Framework for Spatial Data, Proc. IEEE ICDE 2015 (2015).

(2015 年 5 月 14 日受付)

原 隆浩 (正会員) hara@ist.osaka-u.ac.jp

1997 年大阪大学大学院前期課程修了。2004 年より同大学院情報科学研究科准教授。工学博士。2008 年、2009 年本会論文賞受賞。2015 年日本学術振興会賞受賞。データベースシステム、モバイルコンピューティングに関する研究に従事。IEEE、ACM、電子情報通信学会、日本データベース学会各会員。