

社会インフラの革新に貢献する 知識化サービス基盤KaaS

KaaS: Knowledge Service Platform Aimed at Social Infrastructure Innovation

植田 良一

Ueda Ryoichi

佐藤 嘉則

Sato Yoshinori

森 正勝

Mori Masakatsu

中村 浩三

Nakamura Kozo

佐川 暢俊

Sagawa Nobutoshi

日立グループは、産業、交通、電力など社会インフラ分野で発生する大量のデータから有益な情報（知識）を抽出・提供するサービス「KaaS」の実現に向け、研究開発を続けている。

その実現には、膨大な計算能力の確保と、データに付随するプライバシーの保護が課題となる。この課題解決のために、大規模分散処理基盤上に各種分析アルゴリズムを実装可能とするアーキテクチャを開発し、計算リソースの追加だけで処理能力を拡張可能にするとともに、項目同時アクセス制御／ID多重化などの機能を持つ匿名化システムにより、プライバシーを保護するシステムを開発した。現在、開発した技術を保守診断サービスや位置連動サービスなどに適用し、評価・改良を続けている。

今後、異なる業界のデータを組み合わせた分析をも可能とする高い計算能力を持つシステムを構築し、業界の垣根を越えた知の融合を実現する知識化基盤の提供をめざしていく。

1. はじめに

これまで、経営層がITに求める期待の多くは、人手の置き換えによる業務効率向上であったが、近年その期待は、業務効率向上から高付加価値化に徐々に変化しつつある。

一方、IT自体の変化を見ると、RFID（Radio-frequency Identification）やGPS（Global Positioning System）をはじめとする各種センサー技術の発展と普及、センサーからの大量のデータ収集を可能とする通信技術の進展と広帯域通信インフラの整備、大量の計算リソースを安価に供給する仮想化技術やクラウドコンピューティングの進展などが挙げられる。

これらを背景に、日立グループは、GPSやセンサー、ITシステム上に残された各種履歴情報などから収集できる大量の素データを分析して、付加価値の高い情報（知識）を抽出／提供するサービス「KaaS（Knowledge as a Service）」を提案し、その実現をめざして基盤技術の研究を開始した。

ここでは、社会インフラの革新に貢献する知識化サービス基盤KaaSについて述べる。

2. KaaSがもたらす世界

日立グループは、KaaSにより、個々の企業の業務改善だけでなく、社会インフラの高付加価値化、すなわち革新を実現できると考えている。例えば、産業分野では、機器に設置したセンサーからのデータを高度なアルゴリズムで分析することによって当該機器の故障を事前に検知したり、機器の稼働率保証型ビジネスが可能になったりする。また、交通分野では電子乗車券やプローブカー、携帯GPSなどのデータを分析することにより、人々のマクロな流れだけでなく、個々人のミクロな動きまで鮮明に把握／予測できるようになり、この情報を基にしたこれまでにない革新的なサービスや運行制御が実現できると考える。さらに、製造業や流通業の生産・販売計画から予測できる電力消費量と電力会社やコミュニティの発電量など、異なる業界のデータを統合的に解析することにより、広範囲な安定供給、環境負荷低減の観点から動的な需給バランス計画を立案するなどの応用が可能になると考える。

KaaSがめざす社会インフラの革新を実現するには、これら公共性の高い事業会社やサービスプロバイダーなどに蓄積される、個人情報を含む大量のデータを効率的に分析する必要がある。

この要件を満たすために、日立グループは、大きく二つの技術課題に取り組んでいる。一つは、社会インフラシステムに由来する大量のデータからいかに有益な情報（知識）を効率的に抽出するかというITアーキテクチャ上の課題、もう一つは、時によって個人情報や機密情報を含む可能性のあるデータをいかに安全に利活用するかというプライバシー、セキュリティ上の課題である。

2.1 大量データからの知識抽出

社会インフラを支える企業に蓄積されるデータから有益な情報(知識)を抽出するには、処理すべきデータ量が多いことに加え、機械学習など計算量の多い処理が必要とされることを考慮しなければならない。また、抽出した知識を活用して機器の制御を行う場合には高度なリアルタイム性も要求されるため、必要とされる大量の計算リソースの確保と、そのリソースの効率的な活用が課題となる。各分野の代表的なKaaS応用例とその特性を図1に示す。

KaaSは、計算量と即応性の要件から従来の計算機技術では十分な精度が確保できない、コストがかかり過ぎるといった理由によって対応が困難であった応用の実現をめざしている。

2.2 プライバシー保護

実世界の個人の生活にかかわるデータは、「ライフログ」あるいは「パーソナル情報」などと呼ばれ、その活用可能性や保護のあり方に注目が集まっている¹⁾。そもそもライフログなどのデータをどのように扱うべきか、社会的コンセンサスの形成にはまだ一定の時間を要すると思われるものの、少なくとも個人情報保護の観点に立脚したデータ管理が不可欠である。

一方、情報セキュリティ事故の約70%は紛失、誤送信などの過失に起因するものであり、さらに正当なデータ利用者が悪意をもって事故を起こした事例では、一度に100万件以上の個人情報が漏洩(えい)している²⁾。このような事故の様相を踏まえると、技術的にはまず、情報漏洩、不正利用など、リスクの顕在化を未然に防ぐことが目標となる。実現技術としては認証、暗号、アクセス制御などが

活用できる。

さらに、万一の事故発生に備えて、事故の影響を最小限にとどめる手だてをあらかじめ講じることが望ましい。すなわち、ライフログなどのデータの収集、移送、知識抽出、知識活用、廃棄に至るライフサイクルのそれぞれにおいて、できるかぎりデータを安全化しておくという新たなリスク対策が必要となる。ただし、安全化に際しては、知識抽出などの処理を阻害しないように配慮しなければならない。

3. 課題解決へのアプローチ

3.1 大量データからの知識抽出処理技術

通常、KaaSのように、Web経由で提供するサービスは、Web、AP (Application)、DB (Database) の各サーバから成る「Web3階層構造」を持つシステムとして構築されることが多い。しかし、Web3階層構造は、比較的小規模のデータに単純な計算を行うシステムに適しているため、大規模なデータに対して機械学習などの複雑な計算を行うKaaSシステムには適さない。

さらに、KaaSシステムではRFID、GPS、IC (Integrated Circuit) カードなど実世界の人/モノの位置や状態を表す多様なデータを対象として、さまざまな角度からの分析を可能とする必要がある。

日立グループは、「データ量」の要件に対応するために、従来のWeb3階層構造のDBサーバ層に、MapReduce型分散処理方式をベースとするデータ管理層「KaaS-Base」を、また、「データ種」の要件に対応するために、APサーバ層に、多様なデータを統一的/効率的に取り扱うフレームワークをベースとする知識化処理層「KaaS-Core」を、それぞれ配置する構造を採用した(図2参照)。

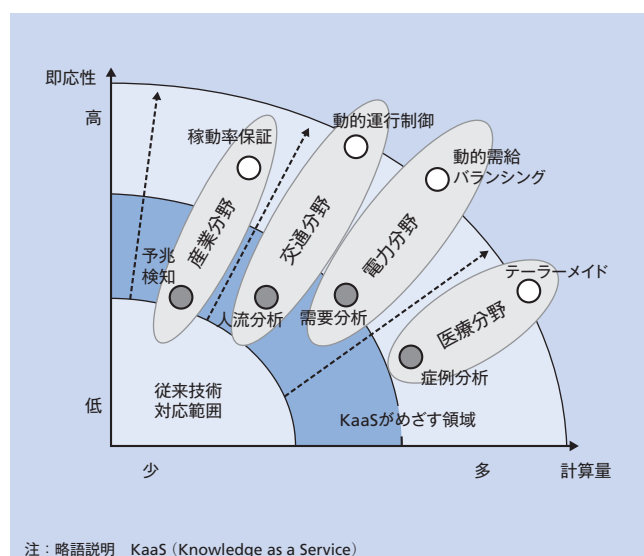


図1 | 産業/交通/電力/医療の各分野でのKaaS応用例とその特性
KaaSは、計算量と即応性の観点から、従来技術では対応が困難であった応用の実現をめざしている。

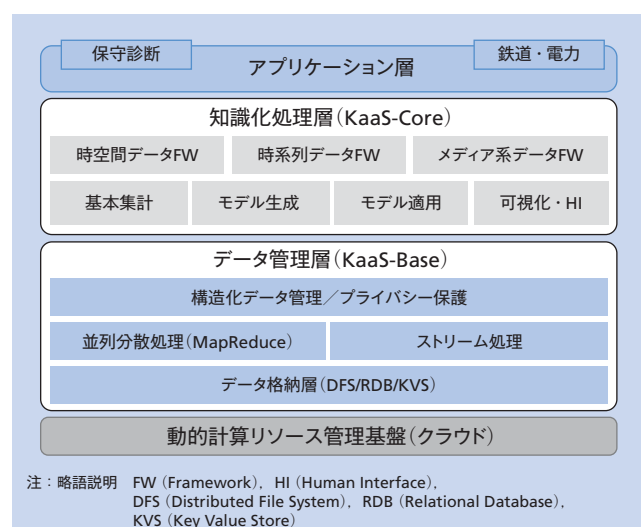


図2 | KaaSシステムアーキテクチャ

MapReduce型分散処理方式をベースとするデータ管理層「KaaS-Base」と、多様なデータを統一的/効率的に取り扱うフレームワークをベースとする知識化処理層「KaaS-Core」から成る。

KaaS-Baseで採用した「MapReduce」はGoogle^{※)}が考案した並列分散処理方式で、プログラムをMapステップとReduceステップの組の集合で記述する。一つのMap処理が他のMap処理に依存しないようにプログラムを記述し、入力データを分割して与えることで、並列分散処理が可能となる。

MapReduce型分散処理方式の採用により、将来のデータ量増加にもプログラムを改変することなく、計算リソースの追加だけで対応できる拡張性の高いシステム構成とすることができる。KaaSシステムでは、MapReduceのオープンソース実装の一つであるHadoopをベースとし、実業応用に対して十分な拡張性を持つように最適化を図っている。また、KaaS-Baseにストリーム処理エンジンを配置し、リアルタイムモデル適用処理を可能とした。

一方、KaaS-Coreは、実業データの型と応用パターンの調査に基づき、分析アルゴリズムなどの必要な部品群と、アプリケーション開発者が共通に利用する機能をデータ種別にライブラリ化したフレームワークの準備を進めている。

アプリケーション開発者が、KaaSシステム上の分析フレームワークや分析部品群を組み合わせることにより、大量のデータから有益な情報を抽出する業務アプリケーションの開発が容易になると考える。

3.2 プライバシー保護基盤

プライバシー保護についての課題を解決する手段としては個人を特定できないようデータを安全化する、いわゆる「匿名化」技術が有効である。図2の「プライバシー保護」における匿名化基盤の概要を図3に示す。

※) Googleは、Google Inc.の登録商標である。

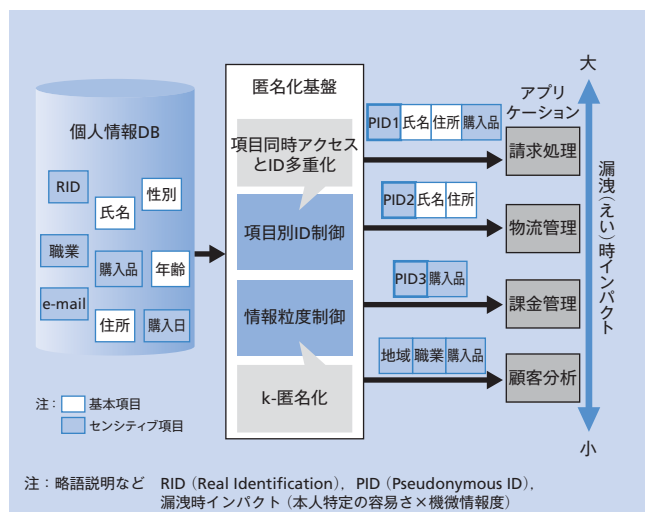


図3 | プライバシー保護基盤の概念

アプリケーションごとに個人ID、利用項目、項目値粒度を制御することにより、利用情報最小化、不正なひも付けの防止を可能にする。

匿名化基盤は、氏名、住所、年齢、性別など、単体もしくはは組み合わせることで個人を特定できる「基本項目」と、それ以外の「センシティブ項目」に分割してライフログなどのデータを管理する。実現する主な機能は、(1) 項目別ID (Identification) 制御、(2) 情報粒度制御である³⁾。

項目別ID制御機能は、基本項目とセンシティブ項目にまたがるフィールドの同時アクセスを制御しつつ、レコードに付与するIDを多重化する。システムの実装形態によって実現手段は異なってくるが、システム設計時に「Privacy by Design」とも呼ぶべき思想に基づいて、同時アクセス制御とID多重化の仕組みを織り込んでおくことが肝要である。

また、KaaSの利用形態においては、「性別」と「住所の一部」と位置情報の履歴を掛け合わせて分析するなど、基本情報を部分的にかつ安全に使いたい場面が想定される。情報粒度制御機能はこのようにときに必要となり、「k-匿名化」がその実現手段となる。k-匿名化は、どのデータについても同じものがk件以上存在するように、データの粒度（あいまいさ）を制御する技術である。日立グループは、粒度制御すべき項目の優先度を指定できるようにするなど、現実的な利用シーンに即したk-匿名化技術の開発を進めている。

以上のように、匿名化はライフログなどに付随するプライバシーの保護手段として一定の役割を果たす。しかし、技術がすべての課題を解決するわけではないことにも留意すべきである。データ収集に際して必要に応じて利用同意を得ることや、利用目的に沿ってデータを使うなど、システムの運用面で社会的要請に応えていくこともまた不可欠である。

4. KaaS適用例

KaaSの適用事例として、保守向け診断サービスについて述べる。

保守向け診断サービスは、保守対象となる機器に温度や圧力などのセンサーを設置し、センサーが収集した状態データを分析することで当該機器の異常を検知するサービスであり、突発故障によるダウンタイムの低減や保守コスト削減を目的としている。

センサーデータの分析は、機械学習技術を適用して、(1) 正常時のセンサーデータの集合から正常と考えられるデータの範囲を特定し、(2) この範囲と現在のセンサーデータとの「ずれ」を計算する、という2ステップから成る。この「ずれ」が機器の特性に応じて事前に設定した閾(しきい)値を超えた場合に「異常」と見なして、部品交換などの保守作業を行うことで、突発的な故障によるダウンタイムを

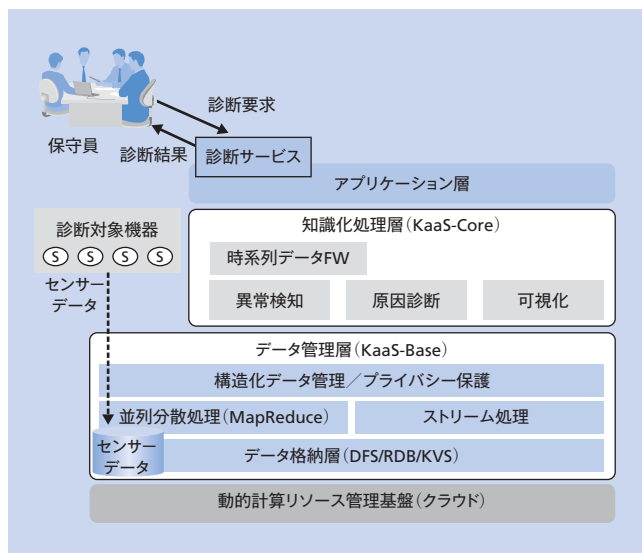


図4 | 保守向け診断サービスの概要

保守対象機器に設置した各種センサーから機器の状態データを収集し、これを分析することで異常の検知、故障時の原因診断などを行う。

低減する。

日立グループは、保守診断サービスアプリケーション、およびそれに必要な分析アルゴリズムをKaaSシステム上に実装し、大量のセンサーデータを高速に分析するシステムを構築中である（図4参照）。

今後、異常時のデータおよびその異常の根本原因、対策方法に関する情報を大量に収集することで、故障の原因分析や交換部品の手配まで、保守業務の広い範囲の効率化に貢献できると考えられる。

KaaSの応用分野としては、上述のような時系列データに関するもののほかに、GPSや電子乗車券ログなど時空間データを基にしたリアルタイム地域人流情報サービスや位置連動サービス、文章や音声などマルチメディア関連データに関する分析サービスなどを検討している。

5. おわりに

ここでは、社会インフラの革新に貢献する知識化サービス基盤KaaSについて述べた。

2010年度からさまざまな実データに対してKaaSの適用を開始した。ここで得られた知見を基に改良を行い、実事業での活用を進める予定である。

また、将来は、電力業界の発電計画と製造・流通業の生産・販売計画とを結び、最適なエネルギー生産・消費を実現するなど、異なる業界のデータを組み合わせた分析を行い、業界の垣根を越えた知の融合を実現する知識化基盤の提供をめざす。日立グループは、これまで鉄道・電力など社会インフラに深くかかわってきた実績と、近年のIT分野での実績を融合することで社会インフラの革新を実現できると期待している。

その革新のために重要なものの一つが知識化サービス基盤KaaSであり、こうした活動が、将来、環境やエネルギーなどのグローバルな課題の解決に貢献できると考えている。

日立グループは、今後も人や地球に配慮した知的創造社会の実現に向けて、KaaSをはじめとする研究開発を推進していく。

参考文献など

- 1) 深層 行動支援サービスに影を落とすプライバシー, 日経コミュニケーション, p.46~50 (2009.2.1)
- 2) NPO日本ネットワークセキュリティ協会: 2008年 情報セキュリティインシデントに関する調査報告書 改訂版1.3 (2009.11), <http://www.jnsa.org/result/2008/surv/incident/index.html>
- 3) 佐藤: 安全・安心社会の実現に向けたプライバシー保護技術の研究動向, 電子情報通信学会第17回バイオメトリックシステムセキュリティ研究発表会 (2009.3)

執筆者紹介



植田 良一

1994年日立製作所入社, システム開発研究所 情報サービス研究センタ uVALUEイノベーション研究部 所属
現在, KaaS, 分散処理技術の研究開発に従事



佐藤 嘉則

1994年日立製作所入社, システム開発研究所 情報サービス研究センタ 第七部 所属
現在, 情報セキュリティの研究開発に従事
情報処理学会会員



森 正勝

1994年日立製作所入社, システム開発研究所 情報プラットフォーム研究センタ 第二部 所属
現在, クラウドコンピューティングの研究開発に従事
情報処理学会会員, 計測自動制御学会会員



中村 浩三

1977年日立製作所入社, 日立研究所 所属
現在, 予防保全システムの研究開発に従事
電子情報通信学会会員



佐川 暢俊

1985年日立製作所入社, システム開発研究所 情報サービス研究センタ 所属
現在, 大規模情報システムの研究開発に従事
情報処理学会会員