

ソーシャル・ビッグデータの利活用によるオープンスマートシティの構築 Creating an Open Smart City based on Social Big Data

徳田 英幸¹ 岩井 将行² 中澤 仁¹ 瀬崎 薫³ 上田 修功⁴ 関本 義秀³

Hideyuki Tokuda Masayuki Iwai Jin Nakazawa Kaoru Sezaki Nonori Ueda Yoshihide Sekimoto

1. はじめに

多様なスマートデバイス、センサなどの進化により、さまざまなデータがリアルタイムに収集、流通、蓄積、解析され、社会的課題を解決し、新しいコネクテッドサービスやアプリケーションを創出するスマートシティプロジェクトが数多く進められている。欧州においては、EUのHorizon 2020における次世代インターネット技術の研究開発の成果であるミドルウェア FI-Ware をスマートシティの枠組みで利活用するプロジェクトや、IoT と Cloud の融合によるスマートシティの構築プロジェクト[1]などが進められている。米国においては、NIST や US-Ignite を中心にIoT や CPS の技術開発をスマートシティやスマートコミュニティの構築にむけて活用する Global City Teams Challenges コンテスト[2]などが活発に開催されている。国内では、総務省における ICT 街づくり推進事業、文部科学省ビッグデータ利活用のための研究開発事業のプロジェクト[3]や国立研究開発法人情報通信研究機構(NICT)によるソーシャル・ビッグデータ利活用・基盤技術の研究開発事業[4]が行われている。

本稿では、ソーシャル・ビッグデータ利活用・基盤技術の研究開発事業プロジェクトの1つである SODA プロジェクトについて、プロジェクトの目的、技術的課題および、いくつかの成果について報告する。

2. SODA プロジェクト

SODA (Social Open bigData) プロジェクトは、NICT のソーシャル・ビッグデータ利活用・基盤技術の研究開発事業の1つで、平成 26 年 6 月から開始した。企業などで利用されているエンタープライズ系のビッグデータではなく、国や地方自治体が公開しているオープンでソーシャルなビッグデータを利活用し、スマートシティを構築していくことを目指している。構成メンバーは、慶應義塾大学、東京大学、東京電機大学、NTT から構成されている。

SODA では、自治体、企業、市民等様々なステークホルダの生成するオープンデータやそれを活用したサービスが、市民の生活の質(QoL)を向上させてくれる街や都市を、“オープンスマートシティ”と呼ぶ。従来のスマートシティで目指されている都市エネルギーや交通渋滞の解消といったシティオフィスの視点からの都市リソースの最適化だけでなく、実空間情報を含むオープンデータの利活用を進めることによって、住民の視点からも、QoL 向上に資するサービスや災害時の安心・安全の向上のためのサービスを容易に実現可能とするソーシャル・ビッグデータ利活用基盤の構築を目指している。さらに、その基盤を利用し都市のイ

ンフラストラクチャや人流などをリアルタイムにマネジメントすること、すなわち“リアルタイム都市マネジメント”を可能とすることを目的としている。

2.1 SODA 基盤ソフトウェアの構成要素

SODA 基盤ソフトウェアは、図 1 に示すように、構成要素を 4 つのレイヤに分割し、研究開発を進めている。最下位層がソーシャルビッグデータの収集レイヤである。従来のセンサハードウェアのように特定の場所に設置されたセンサ群だけでなく、公共車両を始めとする移動体に設置されたセンサ、スマートフォンなどを利用した参加型センシングによるセンサ(Human as a Sensor)、さらに web 上に公開されている静的なデータを仮想センサ化したセンサ(Web as a Sensor)から、データが同一のフォーマットでメタデータとともに収集されている。第 2 層は、センサデータの流通レイヤであり、センサデータの保護、増幅、配送を行っている。第 3 層は、解析機能を提供しており、異種データの統合、都市状況把握、未来推定といった機能を提供している。最上位層は、平時、有事に利用できるリアルタイム都市マネジメントに関するアプリケーション群である。



図 1 SODA 基盤ソフトウェアの概要

3. SODA 基盤ソフトウェア

各レイヤで開発されている要素技術は、図 2 のように表される。以下の節では、各要素技術について概説する。

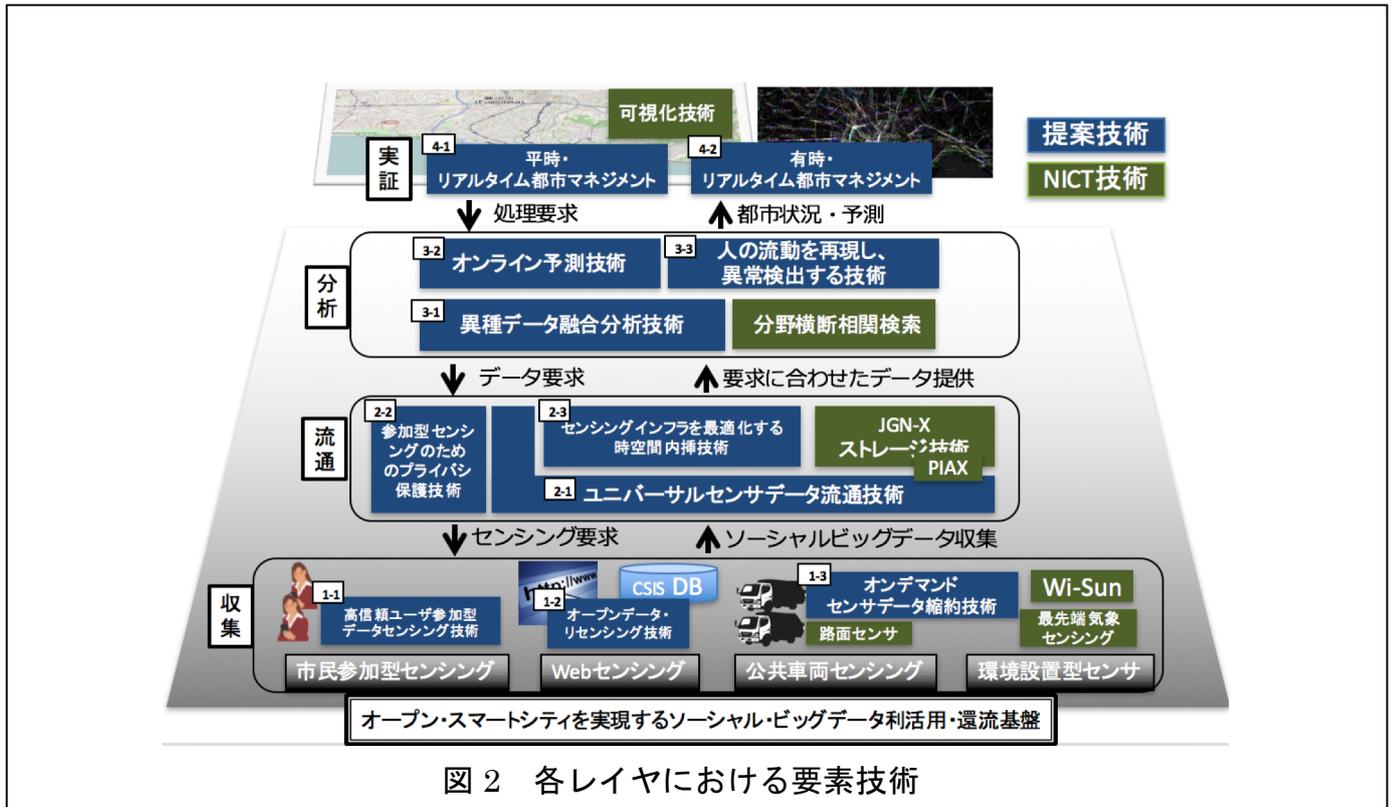
3.1 ソーシャル・ビッグデータ収集技術

1 慶應義塾大学環境情報学部

2 東京電機大学未来科学部情報メディア学科

3 東京大学 空間情報科学研究センター・生産技術研究所

4 NTT コミュニケーション科学基礎研究所



収集のレイヤでは、高信頼ユーザ参加型データ収集技術、オープンデータ・仮想センサ化技術、オンデマンドセンサデータ縮約技術などを実現している。高信頼ユーザ参加型データ収集においては、都市の騒音源や落書き等をはじめとする、人間の五感を用いることで有効に検出する情報の獲得を目的として、スマートフォン等を活用した参加型センシングプラットフォームを構築している。同プラットフォームでは、ユーザの理解度やスキルレベル、センシング時の周辺状況等によってデータに生じるムラを低減するため、ユーザの属性に応じて収集データをクレンジングし、参加型センシングの信頼性を向上している。また、スマートフォン側で測定できる気圧データと解析処理（エッジアナリティクス）を併用し、本来のGPSでは計測できない地下鉄内で移動しているユーザの位置情報なども予測・生成する手法[6]なども提供している。

オープンデータ・仮想センサ化技術は、Web上に公開されているオープンデータを仮想的なセンサに自動/半自動で変換する機能を提供する。大気汚染物質濃度やダム貯水量、幹線道路沿いに設置されたカメラ画像など、国または地方自治体がウェブページ上で定期的に更新するデータをスクレイピングし、後述するソーシャル・ビッグデータ流通技術を通じてストリーム化する点が特徴である。これまで約50万仮想センサを運用し、1日あたり10ギガバイト程度のデータを生成している。

オンデマンドセンサデータ縮約技術は、定期的に変化するセンサデータについて、代表的な変化パターンを記憶させておき、そこから逸脱した場合のみに必要な情報を送ることにより、大幅にセンサデータ量を縮約する。

3.2 ソーシャル・ビッグデータ流通技術

流通のレイヤでは、ソーシャル・ビッグデータを構成するレガシーセンサデータ、ヒューマンセンサデータ、仮想センサデータなど様々なセンサデータを統一かつグローバルに管理し、流通させる機構を提供する。本技術はPublish/Subscribeモデルに基づいて構築されており、これによってデータを生成するセンサ側と、それを消費するアプリケーション側の独立性を高めている。また、ウェブページ (<http://sox.ht.sfc.keio.ac.jp>) では、Java言語やPython言語、Javascript言語等、多様な言語を対象としてAPIを提供し、アプリケーション開発を促進している。

センサデータに対するアプリケーション側からの要求を仮想データソースとして定義し、仮想データソースと実データソースのダイナミックなマッピングを提供し、センサデータの統一的な配送を実現している。データの流通とグローバルなマッピング管理には、XMPPベースの基盤ソフトウェアSoxfire [5]で実現している。同ソフトウェアでは、仮想データソースごとに、データに対するアクセス制御が可能であることから、今後は行政オープンデータだけでなくエンタープライズデータの利活用にも応用可能である。また、後述する第3層においては、仮想データソースの一部として実データソースを複数組み合わせで定義可能とし、データに対する柔軟な分析・処理を可能としている。

また、高精度な固定センサだけでなく様々な精度のハードセンサ、仮想化されたセンサ、ヒューマンセンサが混在する環境において、必要なデータの収集や蓄積を実現するために、センシングデータを時間的、空間的に補完し必要なデータを推定する機能も提供している。

3.3 ソーシャル・ビッグデータにおけるプライバシー保護技術

高信頼ユーザ参加型データ収集のように、人を介してデータを収集する際には、人の位置情報をはじめとするプライバシーが収集データに含まれる可能性があることから、プライバシーを保護する技術が必要である。そこで本プロジェクトでは、人がスマートフォン等の端末から送信するデータにノイズを加えることで秘匿化するデータ摂動 (Perturbation) 技術を提供する。同技術では、個々の値は復元できないが、その分布を統計的に復元可能にする。特に連続値や2者択一問題等、これまでのデータ摂動 (Perturbation) 技術では困難だった値に対するプライバシー保護手法を提供している[8]。

具体的には、モバイル端末側は実際のセンシングデータと一定のルールで異なるデータ (プライバシー保護処理済みデータ) をサーバへ送信する。サーバ側は、多数のモバイル端末から収集したプライバシー保護処理済みデータを集計し、ある数式を用いて統計データを再構築する。このため、サーバ側は個々の端末のセンシングデータを知り得ず、ユーザのプライバシーが保護される。このとき、統計データの再構築精度がモバイル端末側でのプライバシー保護処理済みデータ生成ルールによって影響を受ける。そこで本技術では、ある種のダミーデータを用意し、その選択方法に工夫を施すことで、高精度なデータ再構築を可能としている。

また、Perturbation を利用した参加型センシングを効率化するために、秘匿化されたデータから更に観測が必要なエリアを発見し、インセンティブ付けによるデータ送信者の効率的な配備を支援する、データ復元精度推定手法を開発した[9]。

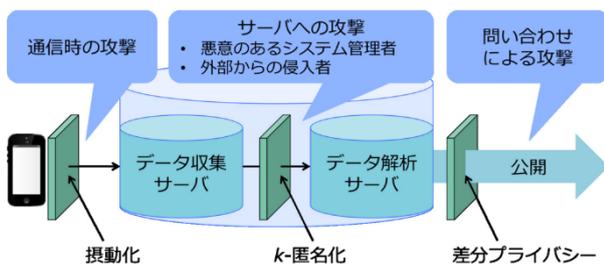


図3 攻撃モデルとプライバシー保護手法

3.4 異種データ融合・未来推定技術

分析のレイヤでは、複数種類の異種データから、異種データに跨る特徴的な潜在パターンを自動抽出するための汎用技術を提供している。従来の異種データ融合技術では、主に、異種データ間の相関関係の分析、すなわち、2つの異種データ間の関係性にとどまっていた。これに対し、本技術では、任意数の異種データを対象に、原理的に全異種データに跨る潜在的な関係性を抽出可能であるという特長を有する。本手法は構造正則化付きの非負値テンソル分解に基づく手法である。テンソルとは、2次元行列を多次元に拡張したもので、異種データに跨る共起関係を表現するのに適したデータ構造である。すなわち、2次元配列では、2種類の情報からなる2項関係 (例えば、ユーザと購入アイテムの関係) しか表現できないが、多次元化 (テンソル

化) することにより多項関係が表現可能となる。非負値テンソル分解とは、テンソルを複数の因子行列に分解する技法である。分解された因子は、元のテンソルを表現する基底故、この因子を異種データに跨る特徴的な潜在パターンと見なすことができる。さらに、構造正則化とは、時空間データの場合、連続する時間、空間ではセンサの値は類似した値になるような制約に相当する。観測データでは、連続する時間、近傍の空間でセンサ値が類似するとは必ずしも限らないが、本構造正則化付き非負値テンソル分解により、観測データをできるだけ忠実に再現しつつ、かつ、構造情報を加味した因子行列が抽出可能となる。

本技術をデンマークのオーフス市内で収集されたスマートシティに関するオープンデータセット CityPulse に適用し、抽出された潜在パターン例を図4に示す。CityPulse データセット[10]には市内の道路419地点で観測された交通流に加え、8箇所の駐車場の車両数、気象データ (風速、気温、気圧、湿度) が含まれる。地図上のオレンジおよび緑の丸はそれぞれ交通流および駐車場の観測地点であり、丸の大きさが抽出された潜在パターン (因子) での各地点のパラメータの強さを表す。図右下のグラフは、各因子が発現しやすい時刻、日付および、気象データのパラメータ

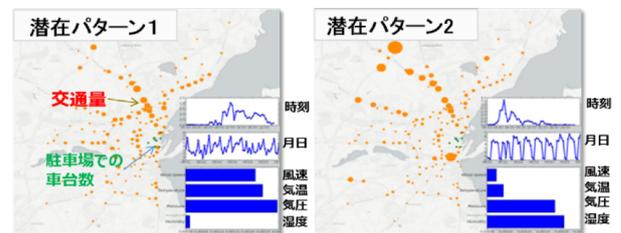


図4 複合非負値テンソル分解による潜在パターン抽出

の強さを示している。本技術により、休日の晴天における交通流と駐車場の混雑発生 (潜在パターン1) や、平日午前中の雨天における交通流と駐車場の混雑発生 (潜在パターン2) といった、都市で計測された異種複合データに跨る特徴的なパターンを自動抽出することが可能である[11]。

また、上記の異種データ融合分析技術に加え、都市においてリアルタイムに計測された多様なセンサデータの時空間発展をモデル化し、近未来をオンライン予測する技術についても研究開発を進めている。

3.5 リアルタイム都市マネジメントサービス

自然災害の発生など有事においても、可能な限りリアルタイムに都市マネジメントを実行するために多様なデータを用いた空間データ解析や視覚化サービスの構築を行っている。ここでは、東京大学空間情報科学研究センター (CSIS) で保有・提供している「人の流れデータセット」[12]といった動的ジオデータをベースに、シミュレーションにより大規模災害時の人の移動パターンを推定・再現し、SNS データと合わせて人々の時空間行動の可視化を行った[13]。また、都市マネジメントの研究対象地域である藤沢市において、過去2~5年分の緊急車両の稼働データを用いることで、時間およびゾーン単位で出勤状況を図5のように可視化した。

これらのデータに基づく時空間的な可視化は、都市で起きている現象や時間変化を俯瞰的に把握できるものであ

る。今後このような基盤データ・公共データに加えて、各種センサデータやソーシャルネットワークデータ等の多種多様なデータをさらに融合・分析することで、行政の意思決定支援はもちろん、市民自身の行動の指針となりえるような情報提供も目指す。



図5 藤沢市における消防・救急車両の過去の出動状況の可視化例（左：消防車，右：救急車）

4. 今後の課題

SODA 基盤ソフトウェアを構成している各層における要素技術を開発/改良していく上では、さまざまな技術的課題がある。収集レイヤでは、スマートデバイスやセンサノード側からのデータに対するクレンジング機能の向上や参加型センシングにおけるインセンティブ問題がある。ゲーミフィケーションや擬人化手法などによる試みが検討されている。流通レイヤでは、フェデレーテッドな Soffire サーバにより、センサデータを流通しているが、リアルタイム性能をあげるための改良が重要である。また、解析レイヤでは、現時点からある時間遡った比較的短期間の過去データのみから予測を必要とする場合がある。かつ、観測データが時々刻々得られる場合、観測データを有効活用すべく、観測と予測を逐次的に実行するオンライン予測技術が必須である。このような予測の枠組みにおいて、予測処理のリアルタイム性能の向上だけでなく、過去データからなる“最適な学習用時空間データ”を導出する課題がある。また、予測技術特有の問題も存在する。例えば、駐車場の満空情報の予測とカーナビゲーションを単純に連携したサービスの場合は、多くの利用者が同一の空いていると予測された駐車スペースを目指して移動してしまう場合がある。このような場合は、駐車リソースの枯渇が発生してしまい、ユーザにとっても、駐車スペースを保有する会社や市にとってもリソースの効率的利用とならない課題がある。

また、技術的な課題ではないが、本質的に整理しなければならない課題がある。本基盤ソフトウェアから収集されるさまざまなデータの“所有権”に関する課題である。書籍などは、著作権により明確に定義されているが、収集されたデータの場合は、所有権が明確に定義されていない。さらに、一次利用、二次利用などを可能とするガイドラインなども整備される必要がある。多くの自治体などが所有するオープンデータの場合、API から利用できるものは、ごく限られた範囲のデータであり、本基盤で利用できる仮想化技術などを利用せざるを得ないのが現状である。

一方、様々なスマートシティに向けたアプリケーションやサービスが構築されているが、どのようにそれらのサービスが持続可能となるか、あるいは、ステークホルダ間どのようにエコシステムを形成できるかといった課題もある。特に、市民を中心に実証されている参加型センシングに基づくサービスの場合は、エコシステムが確立されているものがすくないのが現状である。

5. おわりに

オープンスmartシティでは、自治体、企業、市民等様々なステークホルダの生成するオープンデータとそれらを活用したさまざまなサービスにより、市民のQoLを向上することを目指している。さまざまな実証実験を通じて、要素技術の有効性の検証や持続可能なサービスの社会実装などはさらなる課題であるが、これらの技術的イノベーションだけでなく、ソーシャルイノベーションの創出も重要な課題である。

謝辞

SODA プロジェクトは、NICT のソーシャル・ビッグデータ利活用・基盤技術の研究開発事業の支援によって推進されています。

参考文献

- [1] <http://clout-project.eu/ja/>.
- [2] <https://www.us-ignite.org/globalcityteams/>
- [3] <http://www.info-architects.jp/bigdata/>
- [4] http://www.nict.go.jp/collabo/commission/20140129kobo.html#block_top1
- [5] <http://sox.ht.sfc.keio.ac.jp>
- [6] Satoshi Hyuga, Masaki Ito, Masayuki Iwai and Kaoru Sezaki, "Estimate a User's Location Using Smartphone's Barometer on a Subway," 5th International Workshop on Mobile Entity Localization and Tracking in GPS-less Environments (MELT 2015) co-located with 23rd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL 2015) (2015).
- [7] Jin Nakazawa, Takuro Yonezawa and Hideyuki Tokuda, "Sensorizer: An Architecture for Regenerating Cyber Physical Data Streams from the Web", UbiComp Workshop on Smart Cities: People, Technology and Data (Sept. 2015).
- [8] 皆川昇子, 中澤 仁, 徳田英幸, "H-MDNS: モバイルセンシングにおけるプライバシー保護に配慮した統計データ構築手法", 電子情報通信学会論文誌 B Vol. J98-B No.5 pp.415-424 (2015).
- [9] 鈴木孝男, 伊藤昌毅, 瀬崎薫, "モバイル環境センシングにおける Perturbation 後の復元精度推定手法の提案と評価", 情報処理学会 第 79 回モバイルコンピューティングとパーベイシブシステム研究会(MBL) (2016).
- [10] <http://www.ict-citypulse.eu/>
- [11] 竹内孝, 納谷太, 上田修功, "一般化 KL ダイバージェンスを用いた非負値テンソル補完法と都市交通流データへの応用", 第 8 回データ工学と情報マネジメントに関するフォーラム(DEIM Forum 2016), H7-2 (2016).
- [12] <http://pflow.csis.u-tokyo.ac.jp/>
- [13] 矢部貴大, 関本義秀, 櫻山武浩, 金杉洋, 須藤明人: パーティクルフィルタを用いた災害時におけるリアルタイムな人流推定手法, 交通工学論文集 Vol.2 No.2 pp.19-27 (2016).