

採 択 番 号 : 178B05

課 題 名 : ソーシャル・ビッグデータ利活用・基盤技術の研究開発

個 別 課 題 名 : 課題 B 新たなソーシャル・ビッグデータ利活用・基盤技術の研究開発

副 題 : ソーシャル・ビッグデータ駆動の観光・防災政策決定支援基盤の研究開発

### (1) 研究開発の目的

本研究開発は、Web データ収集・利活用基盤の構築、社会データを活用した科学的分析に基づく合理的な公共サービスの提供、社会データ基盤の持続的運用可能なビジネスモデル開発と、これらの社会実装を目的とする。

観光など地域経済活性化や防災・減災政策などの問題解決の難しさは、部分的でしかも不完全な情報やデータに基づいて、リスクや利益を推定し、主観的判断によって、意思決定を行わなくてはならないことにある。そこで、社会経済分野の多種多量な情報やデータを収集し、科学的分析手法に基づいてデータを解析し、国や自治体の政策決定や企業などの経営の意思決定を支援する ICT システムとサービス開発基盤を構築する。

具体的には、情報空間にアップロードされる様々なデータを収集する社会データ収集基盤を構築する。また、自治体がつもつオープンデータや、Wi-Fi アクセスポイントのログデータを連携させ、俯瞰的に眺めるツールや、各種データの連携・分析によるサービスの合成や政策・意思決定支援を行うソーシャル・ビッグデータ駆動のデータ中心政策決定支援基盤を実現する。

本研究開発で扱う社会データは、国勢調査などの公的統計データ（これは、e-Gov データ、オープンデータと呼ばれる）、宿泊施設、観光施設、公共交通の Web 予約データや賃貸不動産データ（“BOOK-log” と称す）、Wi-Fi アクセスポイントのアクセスログ、天気や台風、地震などの気象データ、地凶データ、食データ(FOOD-log)からなる。これらデータを中心とした科学的分析を行い、人やモノを制御する情報やサービスを合成し、迅速かつタイムリーにフィードバックする技術的・社会的仕組み、および、主に観光客を対象とした行動予測を実現する。特に、本研究開発・実用化では、宿泊施設、観光施設（ゴルフ場）、高速バスの Web 予約データ、賃貸不動産データである BOOK-log および Wi-Fi アクセスログを対象とし、データの収集・蓄積・共有を行うための高度社会データ基盤を構築し、不良設定逆問題解法としてのデータ統合方法を確立し、人やモノを制御する ICT 情報サービスを合成して、迅速かつタイムリーにフィードバックするデータ駆動政策決定支援システムを研究開発・実用化する。

### (2) 研究開発期間

平成 26 年度から平成 29 年度（4 年間）

### (3) 実施機関

大学共同利用機関法人情報・システム研究機構 国立情報学研究所  
大学共同利用機関法人 情報・システム研究機構 統計数理研究所  
国立大学法人山梨大学  
国立大学法人長崎大学

(29-2)

(4) 研究開発予算（契約額）

総額 120 百万円（平成 29 年度 21 百万円）  
※百万円未満切り上げ

(5) 研究開発項目と担当

**課題1 山梨県・八ヶ岳観光圏を対象とした観光政策・意思決定支援基盤の研究開発（山梨大学）**

課題1-1 ペンションなど宿泊施設およびゴルフ場のBOOK-logの収集基盤の構築  
および収集データを利用した稼働率推定手法の確立（山梨大学）

課題1-2 山梨県の路線バスを対象とした人流推定手法および運行計画の最適化手法  
の確立（山梨大学）

課題1-3 路線バスと宿泊施設の連携による観光客分布予報の確立（山梨大学）

課題1-4 八ヶ岳観光圏を対象とした実態調査の実施（山梨大学）

**課題2 長崎県を対象とした観光政策・意思決定支援基盤の研究開発（長崎大学）**

課題2-1 Wi-FiアクセスログおよびGPSデータの有用性の実証とリアルタイム観光  
客分布の可視化システムの構築（長崎大学）

課題2-2 Wi-FiアクセスログおよびGPSデータ、統計データの連携による観光客数  
および分布推定手法の開発（長崎大学、長崎県、国立情報学研究所）

課題2-3 観光船の入港データおよび船内のWi-Fiアクセスログに基づく外国人観光  
客を対象とした観光客予報の確立（長崎大学、国立情報学研究所）

課題2-4 観光客が発信した情報を収集・追加し、観光地の最新情報を発信する“自  
己拡張型オープンデータプラットフォーム”の確立（長崎大学、国立情報学研究  
所）

**課題3 民泊を想定した賃貸不動産の占空状態の推定モデルの研究開発（統計数理  
研究所）**

課題3-1 賃貸不動産情報の収集の効率化（統計数理研究所）

課題3-2 賃貸不動産の占空状態遷移モデルの構築のための実地調査（統計数理研究  
所）

課題3-3 占空状態の促進のための要素の解明（統計数理研究所）

**課題4 観光政策・意思決定支援基盤の共通プラットフォームの開発およびモジュ  
ール化に関する研究開発（国立情報学研究所）**

課題4-1 Web/SNS空間のデータ収集・可視化システムのモジュール化およびAPI  
標準化（国立情報学研究所）

課題4-2 高速バスを対象としたWeb/SNSデータに基づくサービス合成の実証実験  
（国立情報学研究所）

課題4-3 観光政策・意思決定支援の共通プラットフォームの構築（国立情報学研究  
所）

(6) 特許出願、論文発表等

|      |            | 累計（件） | 当該年度（件） |
|------|------------|-------|---------|
| 特許出願 | 国内出願       | 0     | 0       |
|      | 外国出願       | 0     | 0       |
| 外部発表 | 研究論文       | 14    | 2       |
|      | その他研究発表    | 202   | 31      |
|      | プレスリリース・報道 | 14    | 3       |
|      | 展示会        | 9     | 4       |
|      | 標準化提案      | 0     | 0       |

## (7) 具体的な実施内容と成果

## 課題 1

【要約】 観光客を回遊させるための動線の最適化を図るために、バスの停留所ごとのバス乗降数情報を取得するシステムを開発し、データを収集した。また、バスの乗降客数を提供するサイトを構築した。

ハケ岳ツーリズムマネジメント関係者、関係行政機関、ハケ岳方面の路線バス(季節運行)の運行主体、及びやまなしバスコンシェルジュシステムの運営組織との持続的連携協力関係を築いた。

観光地としてのハケ岳地区の現状を調査した結果に基づき、ハケ岳高原や清里地区の観光名所等を巡る周遊バスを対象として、バスの位置情報、バス内の混雑情報などを収集し、データベースとして蓄積し、それをもとにバスの運行情報を可視化するWeb 情報サービスシステムを構築した。Web 情報サービスシステムの表示例を以下に示す。管理者向け表示と一般旅行客向けの表示の例を示す。



図 2. 管理者向け表示例

さらに、蓄積されたデータを分析し、バスの運行遅延の状況等を分析し、運行計画を策定する際のデータを抽出した。その成果の一部を、第11回日本モビリティ・マネジメント会議にて報告した。分析結果の一部を以下に示す。

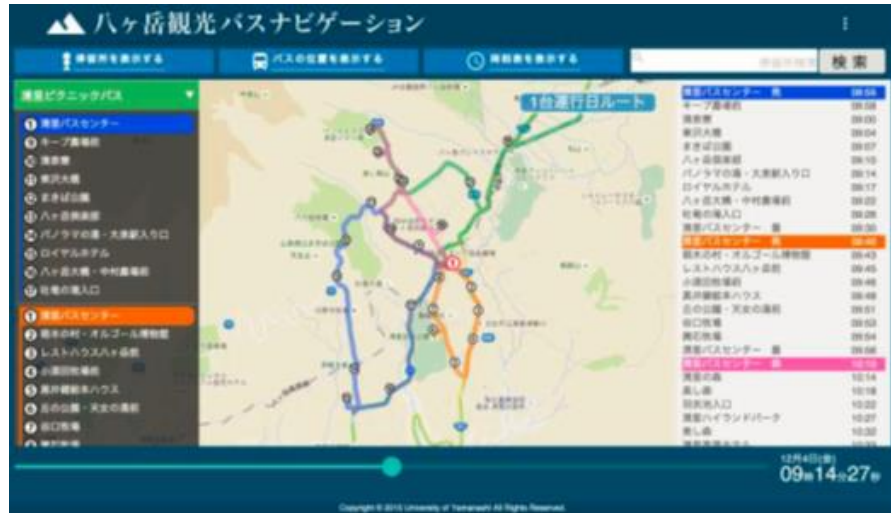


図 3. 一般向け表示例



図 4. 乗降客数の変化の表示例

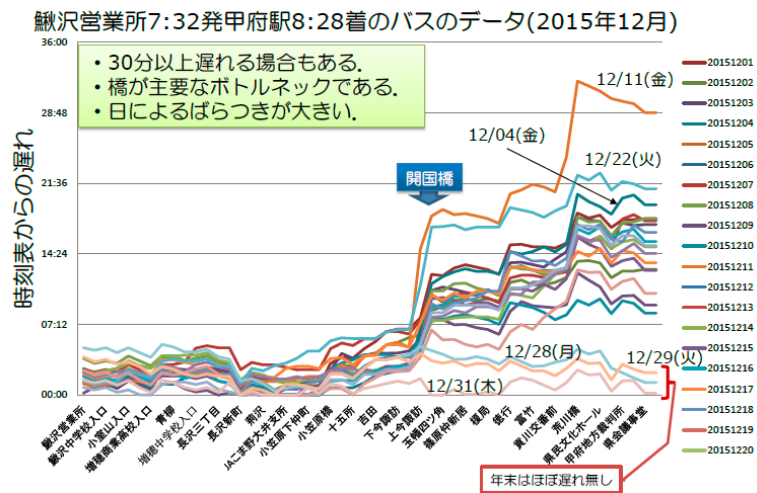


図 5. バスの遅延時間の表示例

また、甲府駅南口に設置された路線バスのデジタルサイネージの設置に向けて、

予備実験を実施した。この実験は、現在、GPSのみで測位されている位置情報の精度を向上させるために、Bluetoothを利用したサブセンサを利用することの効果の検証のために行った。実現可能性について助言、協力し、結果として甲府駅南口にデジタルサイネージの設置に協力できた。

さらに、バスの乗降客数の精度向上を目指して、バスの乗降客数を先の計測方法とは異なる仕組みを構築し、実証実験を実施した。この仕組みは、バスの中にWi-Fiプローブセンサを設置し、乗降者人数を推測する方法に基づく。これらのデータに基づき、人の流れを推定する方法を検討した。これらの研究成果の一部を、第12回日本モビリティ・マネジメント会議及び第22回知能メカトロニクスワークショップ(IMEC2017)にて報告した。分析結果の一部を以下に示す。



図 6. 定時間隔ごとのバスの移動位置の可視化の例

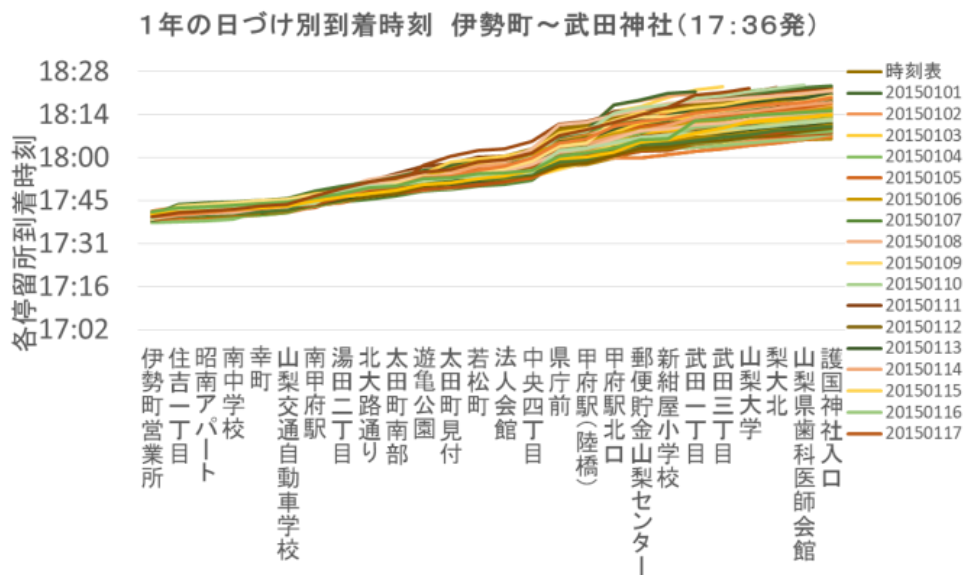


図 7. バスの到着時刻のばらつきの表示例

さらに、八ヶ岳観光圏の2か所にWi-Fiプローブセンサを設置し、観光客等の動きを調査し、人の流れを推定するためのデータを収集した。そのデータから人流を推定するためのデータの可視化の方法を検討した。データの分析結果の一部を以下に示す。

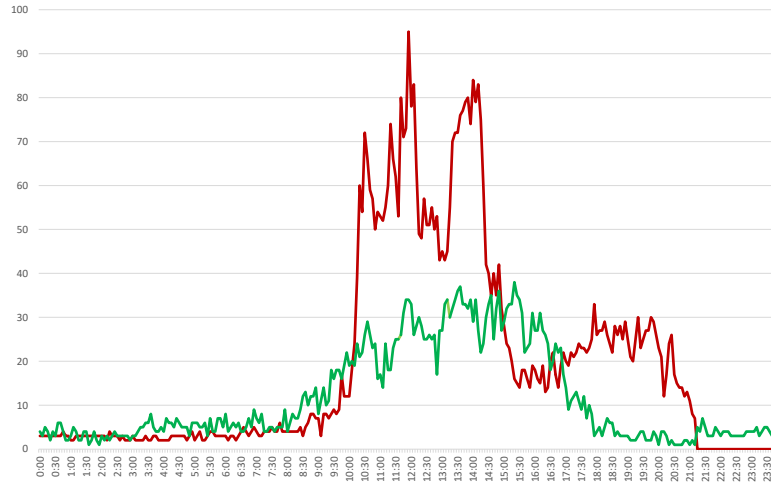


図 8. 萌木の森 ROCK(赤)と道の駅こぶちさわ(緑)へ訪問した端末割合の表示例

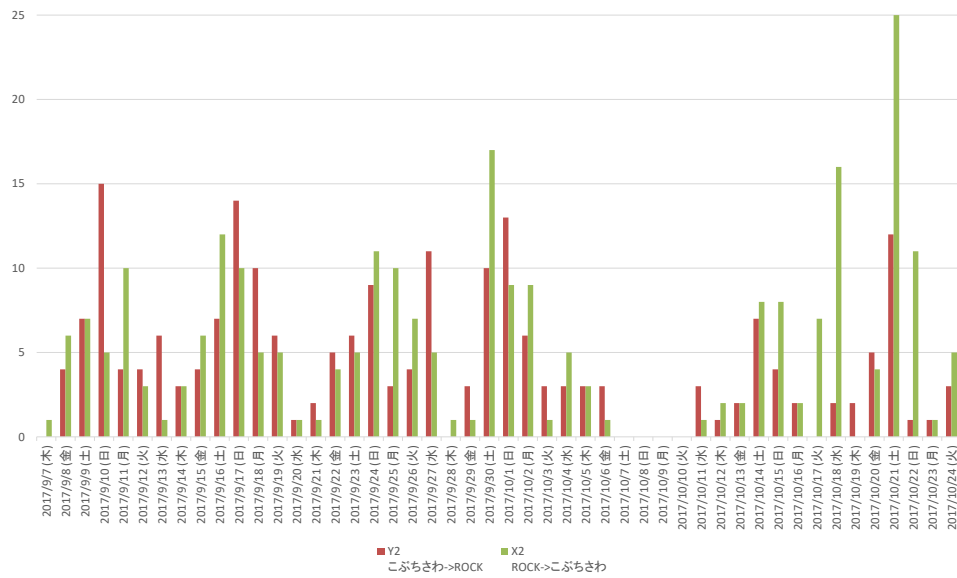


図 9. 萌木の森 ROCK と道の駅こぶちさわ間の移動端末数の表示例  
(赤は道の駅→ROCK を表し、緑は ROCK→道の駅を表す)

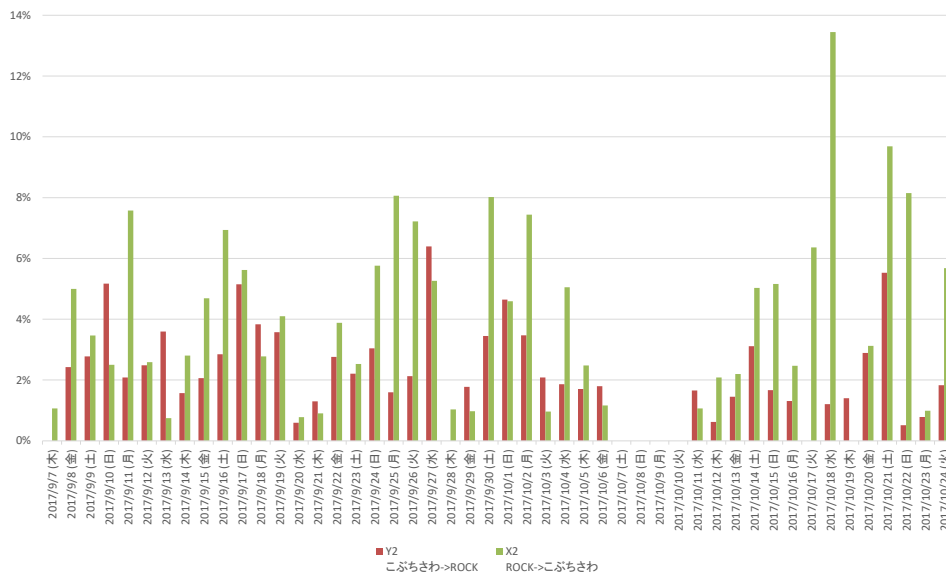


図 10. 萌木の森 ROCK と道の駅こぶちさわに訪問した端末数のうち  
両者間を移動端末数の割合の表示例  
(赤は道の駅→ROCK を表し、緑は ROCK→道の駅を表す)

当初の目標である動線配置の最適化について、行政に提言するまでには至らなかったが、現状を把握するためのデータは提供できていると考えている。また、最適化提案のためのデータも揃っている。

またハヶ岳観光圏を対象とした実証実験のためのシステムの構築し、実証実験を実施し、観光経済活性化のためのデータを取得した。

ハヶ岳観光圏を対象とした実証実験のためのシステムを構築した。やまなしバスコンシェルジュが保有するバス運行データからハヶ岳観光圏のバス情報を抽出し、当該観光圏の観光情報と併せて提示する独立したサイトを構築した。サービスサイトの作成にあたっては、スマートフォン、タブレット端末に適したワンソースマルチデバイス型のUIとして、他の地域でも用いることができるものとした。

また、バスの乗客数情報を用いて、このサイトにて一般ユーザ(観光客)向けに提供した。課題 B-3-3「観光客を回遊させるための動線の最適化」と同時並行で活動した。バスの乗降客数の提供により、ある程度の観光経済の活性化につながるものと考えている。

さらに、蓄積されたデータの効果的な可視化方法について、数種類の表示を検討し、2種類の表示システムを構築した。データ解析のためのデータ表示プログラムも開発した。その研究成果の一部をシンポジウムや学会等で発表した。

## 課題 2

[要約]長崎県内の宿泊施設の稼働率を推定し、また、観光客がどこを周遊しどこを経由しているかについて可視化し情報を提供するための仕組みを構築した。また、自治体が提供する避難所情報、休日当番医情報、バス停情報、車いすの通行可能情報を収集し活用できるプラットフォームを確立し、防災への対応も行った。その成果を長崎県が提供する官民協働クラウドにて、県庁・市役所向けに公開し運用を

長崎大学で行っている。

研究開発項目2 長崎県を対象とした観光政策・意思決定支援基盤の研究開発（長崎大学）

### 研究開発項目 2-1～2-3

ソフトバンクのWi-Fi アクセスポイントのログデータを利用して観光客の動向を把握するために、港湾および佐世保駅にアクセスポイントを設置し観光客の動向について調査を行った。また、提供を受けたデータと観光統計データとの比較を行った結果、利用者に大きな偏りが明らかとなり、これを補正するためのモデルを構築し、推定する手法を確立した。その結果を可視化した例を図 11 に示す。まあ、また、サンキーダイアグラムで県内の流入出量を可視化したものを図 12 に示す。

現状、推定精度が悪く、今後データの蓄積を続けてモデルへのフィードバックを続けていく予定である。現状の達成度は 60%である。

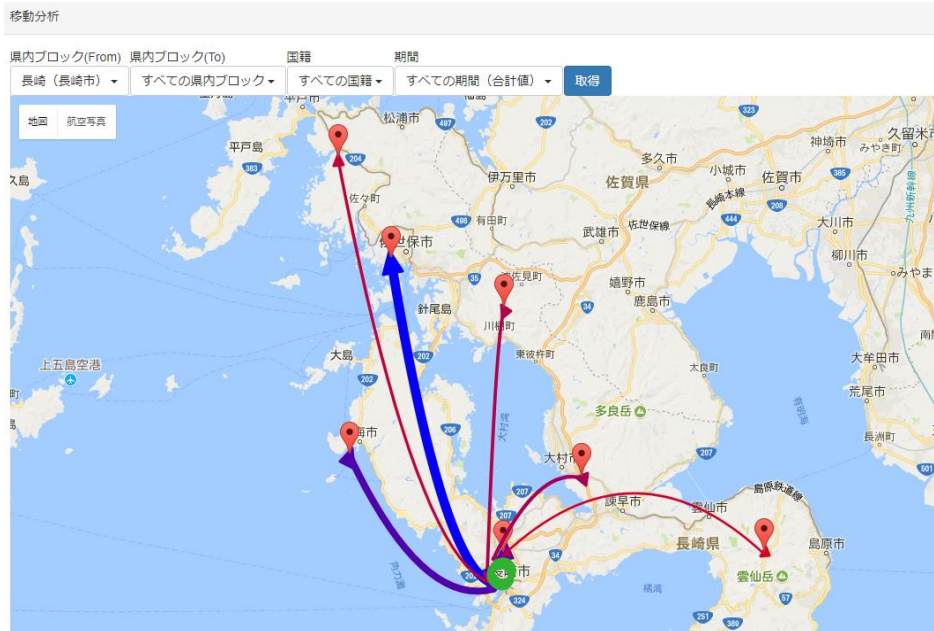


図 11.観光客のブロック間移動の可視化

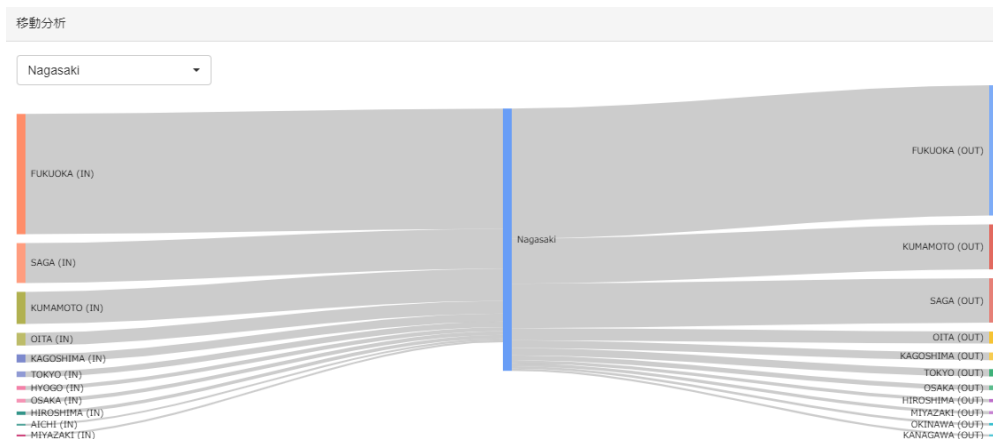


図 12.長崎県内の観光客流入出量のサンキーダイアグラム



### 研究開発項目 2-4 観光客が発信した情報を収集・追加し、観光地の最新情報を発信する“自己拡張型オープンデータプラットフォーム”の確立

ロケーションツーリズムを対象とした「ながさきロケなび」を「長崎さるく」に適用することで自己拡張型オープンデータプラットフォームの拡張性に関する定量的評価を行いプラットフォームとしての有効性を検証した。さらに、「ながさきロケなび」に、自治体が提供する避難所情報や休日当番医の情報、地域社会が提供するバス停情報や地域の車椅子ユーザが発信する情報をバリアフリーストリートビューとして追加することで、防災の観点で強化され、さらに障がい者に優しいプラットフォームを確立した。なお、達成度については、システムとしては完成しているが、社会実装には至っていないため、80%である。

### 研究開発項目 4-3 観光政策・意思決定支援の共通プラットフォームの構築

長崎県の宿泊施設を対象とし、京都市の宿泊施設を対象とした客室稼働率推定手法を適用し、同じように稼働率の推定ができることを明らかとした。また、研究開発項目 2-1～2-3 で確立した Wi-Fi アクセスポイントのログデータによる観光客の動向を可視化するシステムを開発し、図 13 のようなシステムを長崎大学内に構築し実装している。

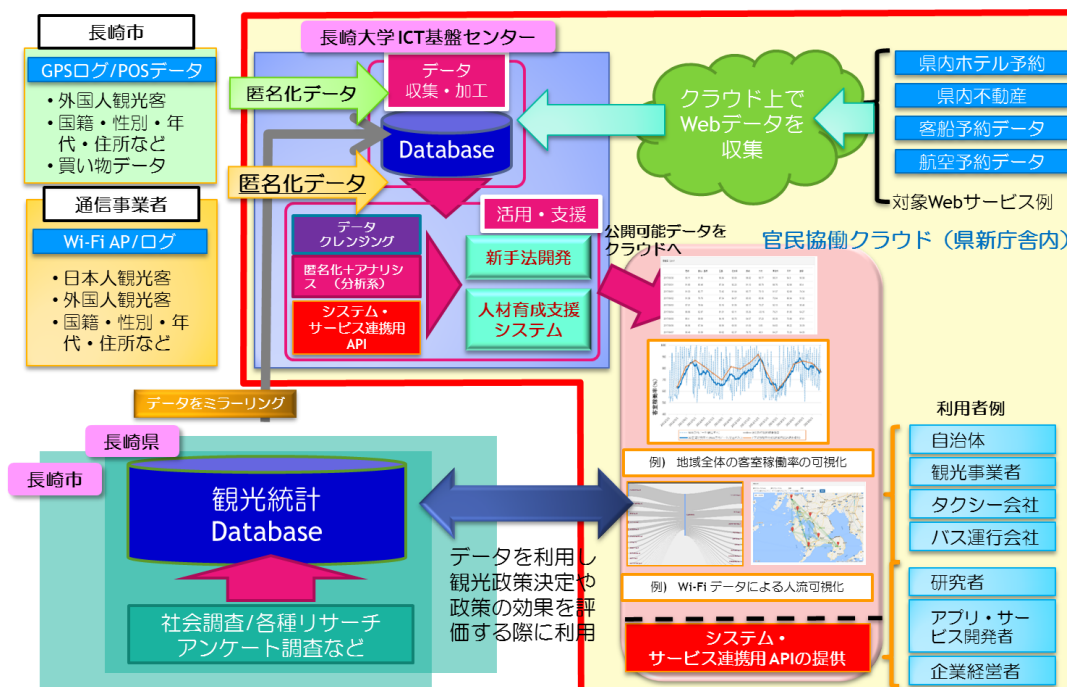


図 13.観光活性化支援システムの概要

このシステムは Web などの情報を大学外で収集し、大学内で実装した開発したシステムを利用して宿泊施設の稼働率や観光客のブロック間の移動を計算し、その結果を長崎県が提供する官民協働クラウド上で可視化している。現在は、長崎

県庁や長崎県内の市役所の観光関係者へ向け公開しており、一般向け公開のために改良すべき点について意見の募集を行っている。自治体向けの社会実装までできているため、達成度は80%である。

#### 課題4

[要約] 本研究開発を通じて、大学共同利用機関ならびに地域の産官学連携を前提とした、観光政策に関するソーシャル・ビッグデータを利活用するためのソフトウェアならびにフォーマット群が構築された。システムを構成する内容として、Web上のデータ収集システムやログデータの構造化に留まらず、そのようなソーシャル・ビッグデータを活用した宿泊予約プランの価格決定支援や購買傾向に関するデータマイニング、ログデータを活用したプライバシー配慮型の群の動態検知などが実現された。

##### ・宿泊予約プランの価格決定支援手法の確立

宿泊プランの価格付けは、宿泊施設経営者にとって重要な問題である。安くすれば客は増えるが安い分収入が減る。反対に高すぎれば客は減り、結果として収入が減る可能性がある。したがって、客数を維持し、収入を増加させるための最適な価格設定があると考えられる。また、客数を増やし、収入を増加させるための価格設定も存在すると考えられる。

そこで、収集したWeb予約データの価格と稼働率を分析することで、稼働率と価格相関について分析を行った。対象を京都市に限定し、価格の上昇率とそれに伴う稼働率の変化について評価を行った。その結果、宿泊施設には価格を大きく変化させないタイプと、需要に応じて価格を大きく変動させるタイプの2タイプに分類できることが明らかとなった。また、それぞれのタイプに対し、稼働率を推定した結果、価格を大きく変化させない宿泊施設は常に稼働率が100%に近い状態となっていた。その一方で、価格を大きく変動させる施設は、価格の上昇と稼働率が反比例していることが確認できた。これらの結果から、得られる売り上げは、同規模の価格・施設数を持つ宿泊施設は、前者と後者のタイプでほぼ同一である可能性が高いことが明らかとなった。

##### ・検索キーワードと宿泊プランの稼働に関する調査および選択されやすいプラン構築手法の検討

宿泊プランをWeb上に提供している宿泊施設は、検索の上位に来るように宿泊プラン名を目立つものにする、空室数を操作するといった工夫をしている。しかし、どのようにプラン名を設定すれば検索されるのかが不明である。事業者の検索キーワードデータベースから、観光に関連することは検索されるタイミングと宿泊施設のプラン名に出現するキーワードの出現タイミングについて検討を行った。その結果、出現タイミングに関する時差は、2ヶ月程度であるという結果が得られた。そこで、検索キーワードとプランに含まれるキーワードの出現頻度、タイミングについて分析を行い、どのタイミングでどのようなキーワードを含むプランの構築をすればよいかを支援するプラン構築支援手法の研究開発を実施し、Web予約サービスの性質を活用したイールドマネジメント手法を検討した。

### ・平均連泊数の推定方法の確立

観光政策決定の科学的根拠となる重要なデータとして「平均連泊数」がある。平均連泊数を伸ばすことで観光地での消費金額が増え、観光地の活性化につながるからである。しかし、この情報もアンケート調査で集計されており、結果が得られるまで時間がかかり、調査頻度を上昇させるとコストもかかる。したがって、Web から収集したデータを利用して、平均連泊数を実時間で把握する手法を研究開発する。宿泊予約データは、残室数が少なくなると（例：10室以下）具体的な数値が表示されない。そのため、直近にならないと、具体的な残室数の変化は把握することができない。

連泊数に関しては、同時に複数の日にまたがって残室数が減少するため、直近の予約の連泊数は把握できるが、それより前の連泊数に関しては把握することができない。そこで、キャンセルに着目する。Web 予約では、キャンセルが発生するとその分残室数が増加する。そこで、キャンセルの数から、残室数の変化が観測できない地点での連泊数の推定ができないかの検討を行い、直近の連泊予約数とキャンセル数を要素とすることで月ごとの平均連泊数を推定するモデルを構築した。宿泊旅行統計調査報告と比較した結果、データが十分に取得できている場合に、精度の高い推定が行えることが明らかになった。サーバーの応答がなくソーシャル・ビッグデータが疎な状態になる場合に、深層学習（LSTM）を活用することで、季節性を考慮した予測が行えることが明らかになり、長崎クラウドをはじめとするソーシャル・ビッグデータ活用時に適用される。

### ・データ収集とタスク管理システム

次に、Web/SNS 空間のデータ収集・可視化システムのモジュール化および API 標準化が行われた。まずデータ収集（クローラー）の実行形態として、ソーシャル・ビッグデータを取得するためのクローラーの実行管理について、これをマスターとスレーブから構成されるクラスター上で実施されるように再設計が行われた（図 14）。マスターは取得するソーシャル・ビッグデータの URI を生成し、スレーブはそれらのジョブを参照、それぞれが実行可能であるジョブを実行、取得されたオブジェクトはリポジトリに送信される。Web 側リソースに対するクローラー割当て状態の一元管理と、ジョブ生成・割当てに関する計画用モジュールの導入が実現される。Web 側のリソースに対して過度なアクセスを実施しないようジョブ実行を制御する一方で、現実世界の事象についてパネル分析を行いたいケースについてはこの制御が欠損値を生じさせ、有意な結果が得られないケースが生じる。このようなケースに対して、ジョブ生成方法を制御することにより分析担当者が実務上で利用可能なデータの収集を実現する

### ・TTPP 手法による可視化と処理の効率化

次に可視化システムについて、リポジトリ側の API を URI として指定することで各種グラフ描画が行える環境の整備を行った。リポジトリ側 API 呼び出し時に集計処理が行われる（または集計済みデータのロードが行われる）が、この集計ステップについて、宿泊施設のスナップショットデータ・バスの遅延状況に関するデータをはじめとするソーシャル・ビッグデータの質ならびに集計内容に応じて処理時間が大きく変化する。本プラットフォーム上でのデータの掛け合わせと

して、Wi-Fi ビッグデータの解析時の知見として得られた TTPP (Tracking Time-window Place Pair) 手法 (図 16) により、宿泊施設のスナップショットデータについても集計処理を効率化させる標準的なデータ構造とすることが可能となった。TTPP 手法は、ある一つのリソースについて時間的な前後関係をもとに2状態から成るテンソルとしてデータ構造を持つ表現系である。宿泊施設のスナップショットデータから施設の稼働率を算出する際には、1階の差分を抽出する必要があった。TTPP 手法を用いることで、この手続きが効率化される。

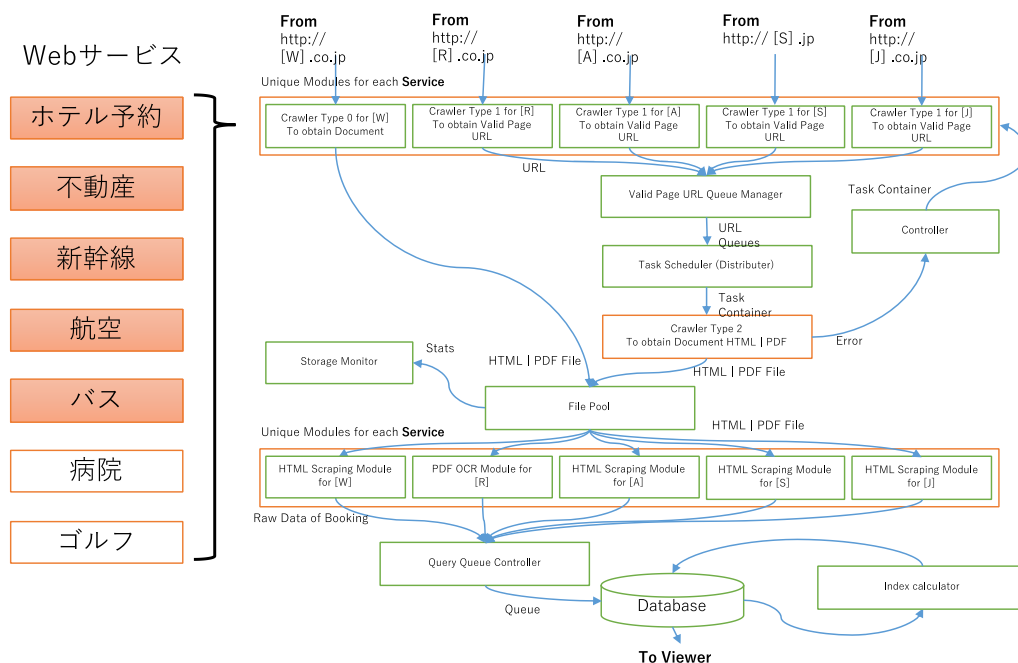


図 14.バックエンドシステムの概要

日時、場所、あるいはそれらの組み合わせの「移動」を示す条件で群流を高速に検索できる。

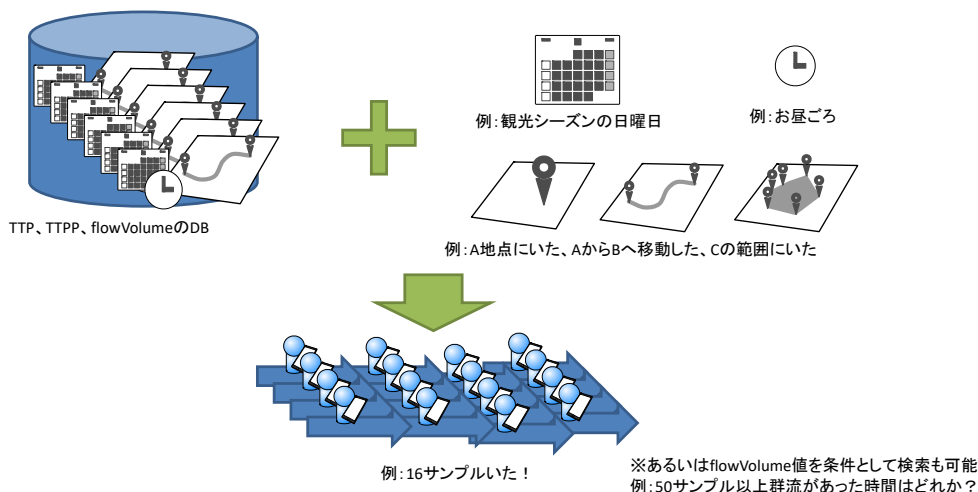


図 15. TTPP 手法による動態サンプル検出

・ソーシャル・ビッグデータを用いた宿泊施設のイールドマネジメント手法のモデル設計

また観光政策に関わるモジュールとして、宿泊施設のイールドマネジメントに関する理論研究が行われた。ペンションに対しては、在庫として少数の部屋しか有さない施設が価格設定を行うにあたって、Web ビッグデータ上で観測される類似する宿泊施設の部屋を一定数ピックアップすることで仮想的に単一の宿泊施設とみなし、各部屋についての在庫の減少状況から自身の宿泊施設の価格設定を行うための最適化モデルを提案した。以下に、ソーシャル・ビッグデータにより1部屋しか有さないペンションを仮想的に統合することで、傾向とばらつきを持つ施設として扱う概念図（図 16）を示す。

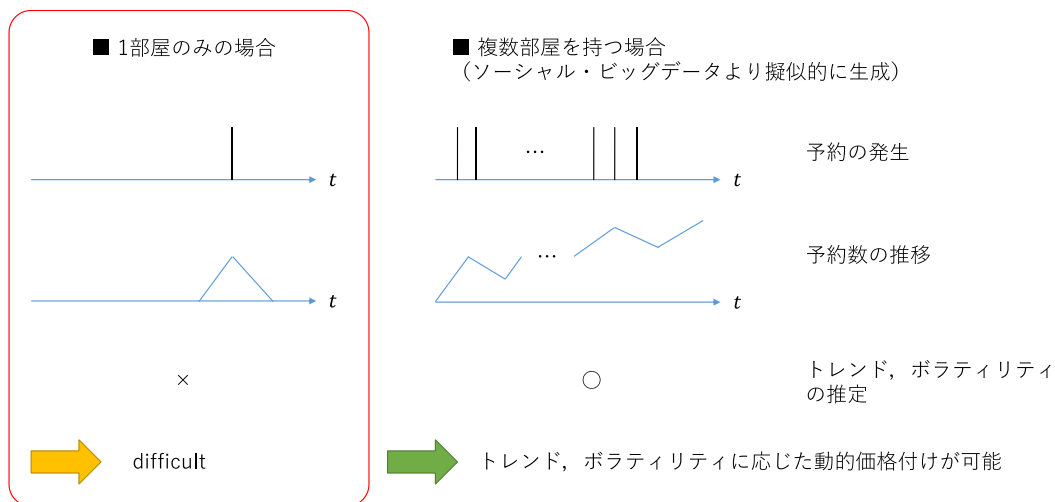


図 16.Web 宿泊予約に関するソーシャル・ビッグデータを用いた擬似的なホテルの生成

次に宿泊施設（ペンション）の経営者に対して、当該エリアの宿泊施設の平均稼働率と競合するエリアでの平均稼働率、ならびにそれぞれのエリアでの平均単価を提供するサービスの開発を行った。課題4-1で整備された API を用いることで、一般の Web サービスから JSON 形式で問合せクエリを送信、レスポンスを受け取ることでサービス上にソーシャル・ビッグデータを集計モジュールで処理した内容を描画する。また宿泊施設の稼働状況に関するアンケート調査を簡便に行えるモジュールの開発を行った。Web ビッグデータを用いた情報提供を行うと同時に、地域の宿泊統計調査を行うための環境（図 17）を整備し、期限を限定して八ヶ岳観光圏によるフィードバックを受けた、これを反映した。



図 17.入出力系を備えた Web インターフェース  
(宿泊施設に関するソーシャル・ビッグデータの可視化を行う)

・研究開発成果の社会実装と持続的な運用

長崎県の長崎クラウドにおける観光事業者向け情報提供サービス上に本プラットフォームの機能組み込みを実現した。具体的には、課題4-2の宿泊施設のオーナー向け Web サービスと同様に自治体クラウド上から本プラットフォームへの API による連携となる。特に長崎県においては、長崎大学において本プラットフォームが運用され、自治体クラウド上で Web ソーシャル・ビッグデータの可視化が行われる。この機能を通じて、長崎県内の観光事業者の情報検索の一助を担う。全国の自治体への展開として、整備されるプラットフォームのモジュールごとの分担について効率化に向けた検討がなされた。

また 2017 年度末には「エビデンスに基づく政策・意思決定を支援／多様なデータを構造化して高速処理／ソーシャル・ビッグデータ駆動の政策決定支援基盤」としてプラットフォームに関するプレスリリースが行われた

前述したデータの収集、蓄積、分析、可視化を行うためのプラットフォームとこれに付帯する共同研究、産官学連携が実施することのできる体制を大学共同利用機関法人 情報・システム研究機構 データサイエンス共同利用基盤施設内に社会データ構造化センターとして展開している。同センターの「ソーシャル・ビッグデータ事業」により持続的な社会実装の推進と根拠データにもとづく政策立案に資する地域の産官学連携の推進を行うことのできる体制を整備した。

(8) 外国の実施機関

該当なし