

NII Today

National Institute of Informatics News

FEATURED TOPIC

CPS — 実社会とサイバー世界をつなぐ

NII Interview

サイバーとリアルを結び、
人々を楽しませる新しい価値を生み出したい

NII Special 1

社会を変えるビッグデータの使い方

That's Collaboration 1

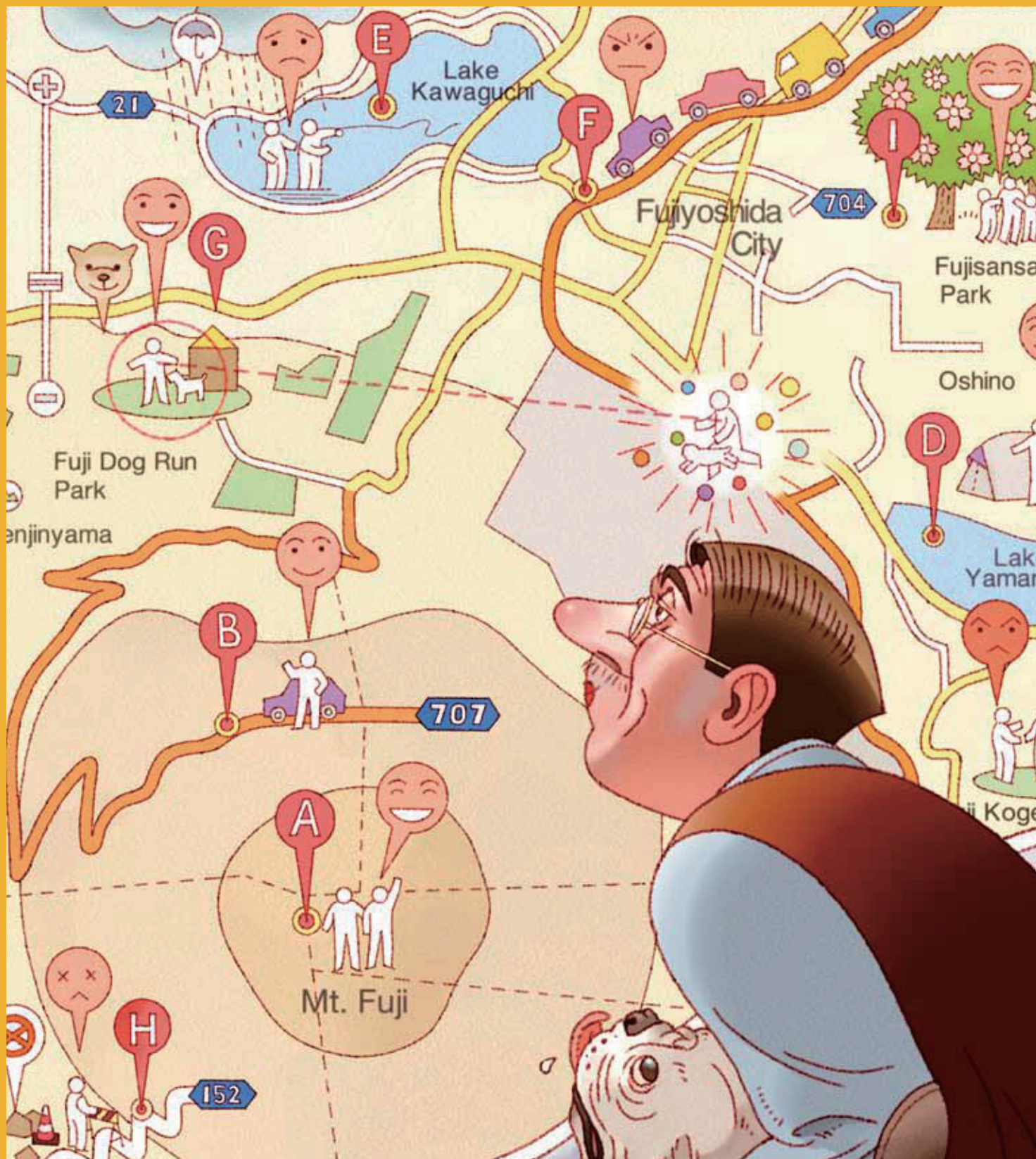
携帯やスマホのデータを観光地の集客に活用

NII Special 2

パーソナルデータ利活用へ向けルール作りが進む

That's Collaboration 2

圧縮・復元技術の最前線



INTERVIEW WITH

相原健郎 Kenro Aihara

国立情報学研究所 コンテンツ科学研究系 准教授
総合研究大学院大学 複合科学研究科
情報学専攻 准教授



サイバーとリアルを結び、人々を楽しませる新しい価値を生み出したい

ショッピングモールでも通勤電車の中からもインターネットを閲覧でき、TwitterやFacebookで経験や思いを共有できる。情報空間は、今や私たちがいつでもどこにいても、つねに寄り添うようにそこにある。実社会と連動しリアルさを増す情報空間をもっと強く現実の世界と結びつけ、その相互作用の中から新しい価値を生み出そうとする研究がある。サイバー・フィジカル・システム（CPS）と呼ばれる分野の先導者の1人、NIIの相原健郎准教授にネットと現実が響き合う未来について聞いた。

滝 サイバー・フィジカル・システム（以下、CPS）は、一般にはなじみの薄い言葉です。

相原 情報学の世界は、第二次世界大戦前にまず計算機科学として始まり、計算機の開発が大きなテーマでした。1990年代にインターネットの時代に入ると、ネット上の情報をいかにやりとりし共有するか、ネットワーク上の情報処理が研究の中心になりました。CPSは次のパラダイムです。Twitterなどを通じて世の中の出来事がネットにも反映すると言われますが、ネットだけでは世の中のことはわかりません。実社会で起きていることをネットの世界に取り込んで解析し、そこに価値を見出して世の中に還元していく。実世界と情報空間、フィジカルとサイバーを統合していくのがCPSです。

滝 あらゆるモノがネットでつながるInternet of Thingsという考え方とはどこが違うのでしょうか。

相原 Internet of ThingsはCPSの実現の仕方の1つではないでしょうか。ただ、すべてのモノがつながっても足りないものがあります。人間です。人間が何を考えて行動し何に価値を見出すのかが入っていません。POSデータの分析からは誰が、いつ、何をかうかがわかって、それは購買という人間の行動の結論に近い部分を捉えているにすぎません。購買に至る過程であれこれ考えるプロセスは捉えられないまま、消えています。私はそこを捉えることが重要だと考えています。「状況センシング」と呼んでいますが、人間の行動のコンテキストにある感情や情動に注目しています。

滝 なぜコンテキストにこだわるのですか。

相原 データやコンテンツはただ存在するだけでは価値がありません。人間の価値観の中で捉えられた時に価値が生まれ情報となります。例えば最高気温が15℃になるというデータはオ

フィスで働く人とお店で商売する人にとって違う意味をもちます。どのコンテンツがどの人にとって価値があるのかは、コンテキストをとらえないとわかりません。情報が生まれる状況、すなわちコンテキストを含めて価値のあるコンテンツになるのです。

もう1つは、情報の受け手が元気だとか、疲れて休みたがっているとか、人が置かれたコンテキストによっても情報の価値は違います。受け手のためにコンテンツを精選する技術が必要です。

滝 人間の心の状況をデータ化するのは容易ではないと思います。

相原 スマートフォンや自動車は今や人の行動を記録するセンサーの固まりです。人に寄り添って行動を把握することができます。私たちは専用のアプリをつくって、ユーザーのスマホに入れてもらい、街での過ごし方を調べました。



現在は東急田園都市線の二子玉川駅周辺（東京都世田谷区）で居住、勤務、あるいは買い物をする人たち数百人に協力していただいています。東急や地元の町内会の方々とも組んで、街でのイベント情報やお知らせをスマホ経由で提供し、利用者の動線だけでなく、どんな情報を目にしてどこに移動したかなど、行動ログを集めています。

さらに2013年度から利用者がスマホの画面を見る際に顔の表情解析を始めています。顔の画像を取得するものではありません。顔の特徴点を見つけてサーバーに送って解析、感情に応じて表れる変化を捉えます。楽しいのか、お疲れなのか、どういう気分なのかを推測できます。街のどこに人が集まっているかだけでなく、集まっている人々の盛り上がりの質まで把握できます。

滝 そうした取り組みを通して、何を達成しようと考えているのですか。

相原 二子玉川などでの試みは街全体の価値をあげることが目標です。街づくりにあたって

人の流れを調べますが、基本的には経験則に基づくところが多いと思います。どこにイベントを打つのが効果的かなど、データに基づき考えることができます。また街を訪れた人に過ごし方を楽しんでもらいたい。どのようなコンテンツを提供すれば行動を誘発し楽しんでもらえるのか。魅力的な街づくりの、いわばOSとして、街の社会基盤の中にインストールしたいと考えています。そのために表情分類の技術などをどう活用できるのか探っています。

滝 NIIではCPS に関してほかにどんな研究に取り組んでいるのですか。

相原 私が携わっているスマホは1つのアプローチですが、ほかに携帯電話会社や自動車会社のデータから世の中の動向を俯瞰する研究も進んでいます。東日本大震災の時に自動車のGPSデータから道路の通行可能状況が推測されたことはよく知られています。車の動きから渋滞の質、例えば自然渋滞が事故なのかなどがわかりますし、気象情報やニュース映像とTwitterの書き込みなどを統合していけば実態

が客観的に見えてきます。そうした状況をもとに利用者は自らの行動を最適化できるようになるでしょう。

いま、多くの人がネットの検索に基づき人気の観光スポットに集中する傾向があり、ランキング上位の場所では大混雑、ほかの観光地は人が集まらないという現象が起きています。検索でバイアスのかかった行動をとるからです。適切な情報提供ができれば、みなが混雑もしないで楽しめるようにできると思います。現在私は、観光庁が観光客の動きを調べるプロジェクトに対し、アドバイスもしています。



インタビューの一言

昨年暮れに長らくNIIの副所長を務められた東倉洋一氏が亡くなられた。東倉さんは人間とモノ・技術とのインターフェースを重視され研究を先導してこられた方だ。記者は東倉さんがNTTやATRの研究所長をされていたころ、マルチメディアという言葉が新鮮だった時代から、情報通信技術の未来についてたくさん教えていただいた。新しい技術を語るときの楽しそうな顔が今も目に浮かぶ。

相原准教授も「物心ともに支援をいただいた。CPSのプロジェクトができたのも東倉先生に引っ張っていただいたおかげ」と話す。外部資金だけでは継続して研究員を雇用できず頭を悩ませていたとき、東倉さんが助け舟を出してくれたという。ご冥福を祈りたい。

滝 順一 Junichi Taki

日本経済新聞社 論説委員
兼 科学技術部編集委員

早稲田大学政治経済学部卒業後、日本経済新聞社入社。産業部、ワシントン支局、大阪本社経済部編集委員、東京本社科学技術部長などを経て、2009年3月から現職。科学技術や環境、医学などを担当する。

“ 情報が生まれる情況、すなわちコンテクストを含めて価値のあるコンテンツになる ”



社会を変えるビッグデータの使い方

多様なデータの統合で見えてきたCPSの具体像と課題

社会システムの稼働情報や、人々のコミュニティ活動情報は、そのほんの一部を切り出してもペタバイトからエクサバイトにのぼる巨大データだ。しかもその形式は多様で、単一データベースに収斂させることすら難しい。センサーネットワークやSNSなどから得られるビッグデータはどう収集・蓄積・分析すれば、実世界に価値あるフィードバックとなりうるのか。ユビキタスサービスに造詣の深いNIIの今井和雄教授と、データ分析に精通する高須淳宏教授に聞いた。



今井和雄 Kazuo Imai
国立情報学研究所
研究戦略室 特任教授

がソーシャルメディアで発信するテキスト情報も有効になるでしょう。交通状況や除雪進捗の把握には、リアルタイムの情報が必要で、タクシーの移動情報の他、除雪車からも例えばスマートフォンを利用して稼働情報を取得したいと考えています。これらのリアルタイムに発生するデータ以外に、過去の交通渋滞データや除排雪記録、交通事故データ、地図データなども状況の知識化のために必要です。全部の情報を集約し、分析・シミュレーション・可視化などの処理を行った結果、最も効果的な除排雪作業指示（意思決定）に結びつけることができ、バスや緊急車両のスムーズな運行や、渋滞

にわたることです。各種のセンサーによる実世界の物理的な状況を示すデータに加え、映像や、TwitterなどのSNSを通じて人が発するテキスト情報も含まれます。

CPSでは、センサーデータに代表されるように、物理世界から逐次情報が得られますので、従来のようにまずはデータをどこかに貯めてじっくり料理するという方法は適しません。データ1つひとつのサイズは小さいとはいえ、膨大な数量に及ぶため、1箇所への集約はもちろん、分散しての蓄積も容易ではないのです。データ圧縮技術のさらなる改良に加えて、データの取捨選択の仕組みが必要になると考えています。

またデータ表現がそれぞれ異なり、統合が難しいことも問題です。以前、銀行などの企業合併の際のシステム統合にあたって、管理しているデータの「名寄せ」が問題になったことがあります。統合するシステムはお互い同じようなデータを扱っているはずですが、その表現がそれぞれ違うので、精度よく統合しようとする表現の違いをうまく吸収して情報を結びつける技術が必要になります。センサーデータの場合も同様の問題が生じます。CPSでは、この問題をデータストリームという、より厳しい状況下で解くことが求められます。

CPSに関わるデータでは、まず、データ発生の「時間」と「場所」をカギにした統合を行います。例えば地図情報とスマートフォンや自動車のGPS情報、SNSに投稿された地名情報などといったデータに絞りこみ、それらを時間軸で組み合わせれば、複数のデータから得られる情報を相互に補完することにより、交通状況をきめ細かに把握できるわけです。

巨大データを分析するには、いつ、どこで、どういうイベントがあったかという切り口でデータを抽出するのが得策です。抽出データに過去の同種イベントの発生記録などを突き合わせると、例えば「A地点で事故が起こると周辺

CPSは現実世界をどう変えるのか？

—CPS（サイバー・フィジカル・システム）が私たちの暮らす現実世界をどう変えるのか、具体的イメージをお聞かせください。

今井 わかりやすい例として、NIIが北海道大学や大阪大学、九州大学と共同で実施しているCPSのプロジェクトにおいて、北海道大学が中心となって計画している「スマート除排雪実証実験」があります。この実験が目指すのは、積雪による道路渋滞を緩和するために最適な除排雪車オペレーションを行うことです。これを達成するためには、さまざまなデータが必要です。積雪状況の把握には、気象データはもちろんのこと、路面状況の計測器からのデータや歩行者

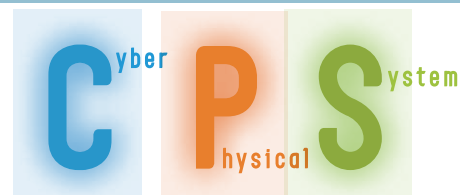
しがちな道路の平均走行速度の向上、また凍結路面の情報発信など事故防止にも役立てられることが期待できます。

膨大で複雑なデータの分析がうまくできれば、例えば公共空間の人の数や位置、動きによって地下街やビル群の空調機器をきめ細かく制御するエネルギー最適利用システムの提供や、災害時には同様の情報を利用して、地下街からの適切な避難経路をそれぞれの個人に提示するなど、さまざまな応用が考えられます。

データ収集・集約・蓄積にまつわる課題は？

—CPSの前提になるデータの収集、集約、蓄積の方法や課題について教えてください。

高須 CPSの課題の1つは扱うデータが多岐



道路状況はこうなる」という予測も可能になるでしょう。NIIの研究プロジェクトでは、自動車のGPSデータから道路の平常時の状況を把握するとともに、平常時と異なる挙動をする自動車を見つけることによって、道路の異常状態をリアルタイムで検知する方法を研究しています。現在は、自動車のGPSデータのみを用いていますが、テレビ放送のニュースやSNS情報の分析を行う研究グループと共同して、情報を多角的に捉える検討も進めたいと思っています。

今井 別の観点からの課題として、プライバシーをどう保護するかという議論も必要です。個人の所在場所、移動状況、SNSへの投稿内容などをどこまで取得してよいのか、また保管や利用の範囲をどうするかについて、社会的な合意が必要になります。

データ管理・分析の方法と課題は？

— データ管理・分析の方法と課題を教えてください。

高須 分析対象のデータには、蓄積された過去データのような「ストレージ系」データ^{※1}と、絶え間なく発生する「ストリーム系」データ^{※2}とがあります。

ストレージ系データは従来からのデータベー

ス管理システム、データウェアハウス、統計解析システムの技術を利用して蓄積・検索・分析されてきましたが、CPSが対象とする膨大なデータを相手に高度な分析を行うには処理性能面でさらに改良が必要です。また、データの背景にある潜在的な情報を抽出するためには、機械学習アルゴリズムの洗練も必要です。私たちの研究グループでは、潜在トピックモデルと呼ばれる統計モデルを用いたCPSデータの分析アルゴリズムの開発に取り組んでいます。

ストリーム系データでは、時間と計算資源の厳しい制約の中で分析を行う必要があります。そのため、いちいちデータベースにデータを格納せずにメモリ上で分析処理し、短時間で結果を出すCEP（複合イベント処理）技術が開発されています。一方で、データストリームの分析では、対象データの範囲を絞ったり広げたりしながら探索的に分

析を行うことも必要になります。現在、データストリームの分析技術と分析のニーズの間にはまだまだ大きな技術的ギャップがあります。このギャップを埋めることは、今後の大きな課題です。

高須淳宏 Atsuhiko Takasu

国立情報学研究所
コンテンツ科学研究系 教授
総合研究大学院大学
複合科学研究科
情報学専攻 専攻長
教授



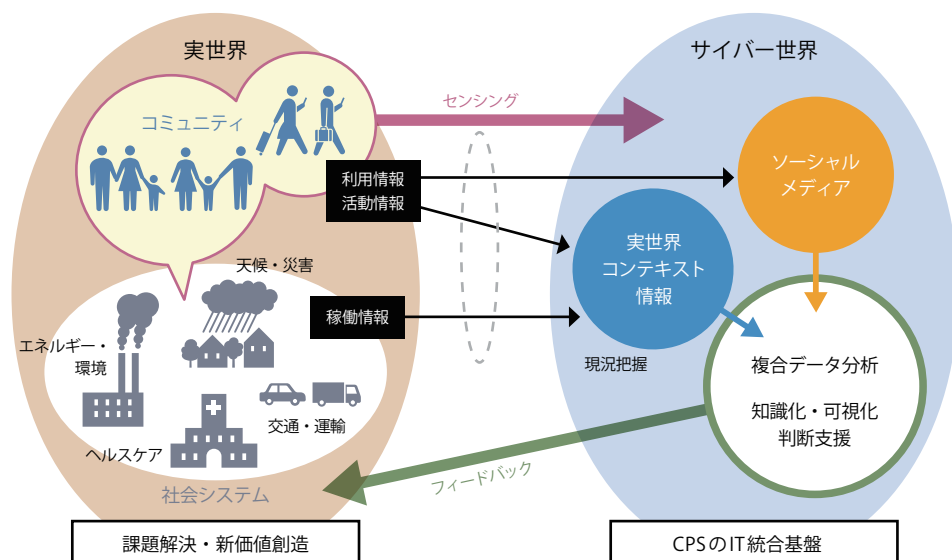
今井 CPSは明快な目的を据えて取り組むことが大切です。その目的に沿って、必要なデータが何かを明らかにし、データ保護を含む管理の条件を洗い出していくというように、総合的に考える必要があります。NIIではここで述べたような課題を解決しつつ、社会・公共・災害対応といった領域に役立つCPS基盤を追究していきます。

(取材・文＝土肥正弘)

※1 一般的な業務システムの処理やIT機器のログ管理、SNSの一括検索などのように、データの発生から分析までの間に時間差があってもかまわない用途に使われるデータのこと。貯めておいた状態で分析処理を行う。

※2 何らかの状態変化に即応した機器制御や、リアルタイムな意思決定支援を行う用途に使われるデータのこと。発生時点で即座に分析処理され、貯められることなく流れすぎていく「フロー型」データ。

(実世界の状況をセンサー等によりリアルタイムに把握し、
データ分析に基づくフィードバックにより社会システムの課題を解決し新たな価値を創造する)



図：ソーシャル・サイバー・フィジカル・システム (CPS)

携帯やスマホのデータを 観光地の集客に活用

観光客の動きを知り、地域づくりに役立てる

観光庁では現在、携帯電話等のGPS（全地球測位網）を用いて収集された約70万人の行動データの集積結果を元に、それらビッグデータを分析することで、新たな観光ルートや観光スポットの発掘に生かそうという研究を進めている。観光庁の川瀧弘之課長、佐世保市観光物産振興局の森永博昭局長、この研究の作業部会座長を務めるNIIの相原健郎准教授に、研究の目的や手法、今後の展開等について伺った。

人の「動き」の情報が 新たな観光地を生み出す

相原 観光庁が進めている観光行動の研究は、携帯電話会社が利用者の許諾を得て収集した個人情報を含まない70万人分のGPSデータから取りまとめられたデータを利用し、観光客の出発地域、移動経路、滞在時間、宿泊の有無等を分析して自治体等に提供、各自治体で活用してもらうというものです。この研究には、どのような目的があるのでしょうか。

川瀧 近年、国内の観光客は減少傾向にありますが、その一方で外国人観光客は昨年1000万人を突破し、過去最高となりました。また、団体旅行から個人旅行へのシフト、ニーズの多様化も進み、観光産業は量的にも質的にも大きく変化しています。観光庁では周遊型観光から滞在交流型観光への転換により、観光産業の活性化を図ろうとしています。そのための戦略づ

くりには詳細なニーズ調査や市場分析が必要です。これまでのような経験則ではなく、客観的なデータを活用して観光客の動きを捉え、的確に分析するために、今回の調査研究を始めました。

調査は現在、8地域で行っています。北海道富良野市、長崎県佐世保市等、複数の自治体が連携することで多くの観光客の誘致が期待される6つの観光圏と、福島県および富士山周辺が対象です。

相原 データを分析すれば個人旅行でも移動経路がわかりますし、意外な人気スポットが発見できる可能性もありますね。新たな観光ルートが発掘できれば、観光マップに加えるなどして観光客の誘致につなげられるでしょう。

森永さん、自治体の観光振興担当の立場としては、この研究をどう捉えていますか。

森永 観光客の周遊ルートや満足度を知るため、これまで佐世保市では年間1万件を目標に14スポットで聞き取り調査をしてきましたが、この方法はコストがかかるうえ、データを解析するのも難しいものでした。今回の調査はマン

パワーをGPSに置き換えて、安く、簡単に、かつ、これまで捉えきれなかった情報も捉えるもので、非常に期待しています。

データによる現実把握は マーケティングに有効

相原 重視するデータはどのようなものでしょうか。

森永 当市では2大観光地であるハウステンボスと九十九島の間をいかに周遊してもらうかが課題であり、戦略づくりにあたっては、観光客はどこから来てどこへ行くのか、昼食はどこで食べているのかなど、人の動きをおさえることが重要だと考えています。移動の方向の情報は観光案内標識等の掲示場所にも関係しますし、ターゲットが明確になることで、効率的にPRや営業活動等ができるようになると思います。

相原 インターネットなどで目につきやすい観光地に観光客が集中すると、周辺地域に影響が



川瀧弘之 Hiroyuki Kawataki
観光庁 観光地域振興課 課長



森永博昭 Hiroaki Morinaga
佐世保市観光物産振興局 局長



図：観光庁が進めるビッグデータを活用した観光行動研究の流れ

出るとい説もありますが。

森永 大きな観光地ができた場合、そこに観光客が集中して周辺には人が行かなくなるという説と、巨大観光地が新規の客を引き寄せ、周囲にも良い影響が出るという説があります。実際どうであるかは、データを分析すればはっきりするでしょう。もし周辺の観光客が減ったとしても、その動向が早くわかれば次の策が打ちやすくなります。

また、本市としては連泊客の動向が知りたいですね。それを分析して満足度の高い食事や魅力的なルートの情報を提供し、宿泊数増加につなげたいと思います。

川瀧 状況を可視化することが重要だということは、どの地域も感じています。明確なデータは、自治体が説明責任を果たす意味でも、民間企業のマーケティングのうえでも役立ち、観光業界全体に有意義だと思います。

森永 行動経路の把握は、「100人の観光客が周った一番人気のコース」のような、新たなお勧めコースづくりにも役立ちます。従来コースと併記すれば、おそらくほとんどが「100人」のコースを選ぶでしょう。そこから観光産業を活性化させる新たな方法を考えることもできます。

データの掛け合わせで行動の意味づけも可能に

相原 ビッグデータの活用で、今後も多様な展開が考えられます。たとえば、GPSのデータ

をTwitterのように特定のトピックやそれに対する態度などが含まれるコンテンツと掛け合わせれば、観光客が何に興味をもって動いているかなど、行動の意味づけができるようになります。また、車やバス、タクシー等の交通機関の動体データと重ね合わせれば移動の状況などもわかるようになります。将来的には、観光客を迎える地元の事業者がリアルタイムに需要に応じて対応できるようなサポートの仕組みもつくられるでしょう。

ところで、現在のデータは観光客だけを抽出していますが、出張のついでに短時間だけ観光するケースも多いと思います。そのような層の情報は活用できますか。

森永 「ついで観光」をする層には期待しているので、データが抽出できれば、アプローチ方法の検討や、ニーズに合わせた体験型プログラムの開発などに活用したいですね。これに限らず、ビッグデータは多様化するニーズに合わせたプログラムづくりにつながると思います。

ITを活用するマネジメントは観光産業にとって大きなステップとなります。しかし、データがあっても、それを生かせなければ意味がありません。データをどう活用し、どう自分たちの観光地づくりをしていくかが鍵となるでしょう。

また、現在のデータには個人の属性は含まれていませんが、マーケティングにおいて年齢、性別は基本情報であり、「40歳代の女性がよく旅行するという説は正しいのか」といったことを、本当は検証したいところです。将来的に高度なフィルタリング機能が実現でき、個人情報に配慮したかたちで属性情報が得られれば、さらに



相原健郎 Kenro Aihara

国立情報学研究所 コンテンツ科学研究系 准教授
総合研究大学院大学 複合科学研究科 情報学専攻 准教授

マーケティングの質が高められると思います。

川瀧 今後は収集データを用いて各地域のニーズを掘り起こすと同時に、他の地域でも展開し、他に必要なデータを検討していきます。外国人観光客の行動については、スマホに観光関連アプリをダウンロードしてもらうなどして情報収集することも考えられるでしょう。

相原 観光アプリといえば、私は愛媛・松山で実証実験を行う中で、旅のサポート・サービスアプリ「ひめパス」※の開発に関わりました。アプリをダウンロードすると街の情報を受け取れるほか、アプリから周遊パスを購入すれば定額で関連施設に入り放題となる、というものです。利用者にはもちろんのこと、地域事業者にとってもメリットがあるアプリであれば、提供しやすいと思います。

川瀧 いずれにせよGPSデータの分析だけでは不十分なので、掛け合わせの手法も合わせて検討し、観光産業のどの部分で活用すると効果的かを具体的に探るなどして、ビッグデータを観光の振興に活用していきたいですね。

(取材・文＝桜井裕子)

※ひめパス

愛媛県松山市周辺の観光エリアや美術館・記念館・博物館などを広く楽しんでもらうためのアプリ。スマホにアプリをダウンロードすると、商店街の約20店舗に設置された「メーカー」のNFCタグやQRコードを読み取るだけで、その利用者の状況に合わせた街の情報を受け取れたり、施設が利用できたりする。URL: <http://himepass.jp/> (2014年3月現在、利用できるパスの販売はしていません。)

パーソナルデータ利活用へ向け ルール作りが進む

プライバシー保護とデータの 有用性を両立させる

CPSが普及すれば、そこには膨大な量のパーソナルデータが集まる。このような技術の進化を背景に、パーソナルデータの利活用とプライバシー保護を両立させるルール作りが、今まさに急ピッチで進められている。個人情報保護法の改正に向けてパーソナルデータの「匿名化」などに関して検討した「技術検討ワーキンググループ」の主査を務めるNIIの佐藤一郎教授に、今後のパーソナルデータ利活用とプライバシー保護の考え方について聞いた。



佐藤一郎 Ichiro Satoh

国立情報学研究所 アーキテクチャ科学研究系 教授
総合研究大学院大学 複合科学研究科 情報学専攻 教授

パーソナルデータ利活用と プライバシー保護はセット

CPS(サイバー・フィジカル・システム)が台頭すると、情報の世界と私たちが暮らす実世界とが、より密接に結びつくようになる。すでに、その前兆といえる動きが始まっている。

例えば、GPS(全地球測位網)などを利用したカーナビや道案内サービスは自動車や人の位置を捕捉する。その情報は利用者だけでなく、社会インフラの効率化からマーケティングまで広く利用できるだろう。実際、CPSによるパー

ソナルデータの収集・分析は、公的機関はもちろん、民間企業にとっても有用だ。例えば、乗車履歴から人々の日々の活動を細かく把握できれば、鉄道運行の円滑化に役立てられることはもちろん、路線周辺の店舗向けのマーケティングにも有用だ。

ただし、CPSが扱う現実世界には人も含まれ、その情報は利用者のプライバシーを垣間見せてしまう。街頭に置いた監視カメラで通行人の「顔」を含む映像を撮影、記録していれば、そこには個人の識別や特定に結びつくデータが蓄積されていく。医療現場の診断記録も、個人にとって重大な秘密と結びついたデータだ。し

たがって今後は、CPSを含む新しい技術を前提にしたプライバシーの保護とパーソナルデータの利活用の両立が重要となる。

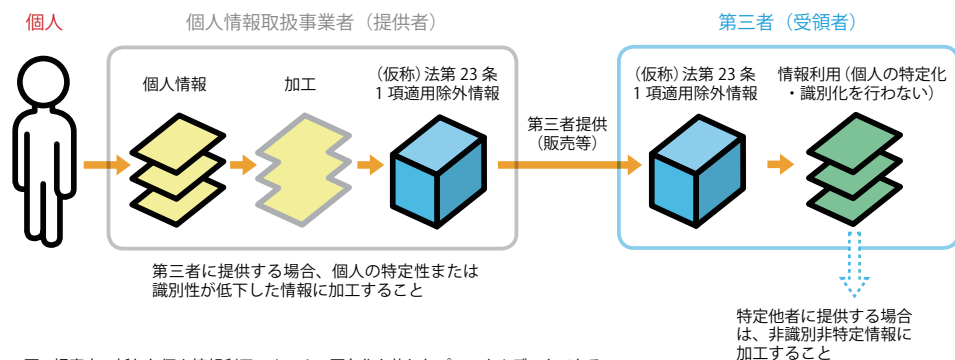
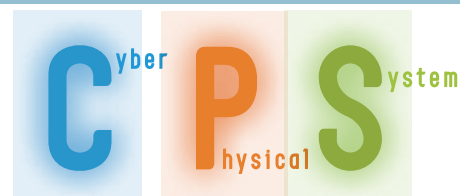
そうしたなかで現在、2003年に成立した「個人情報の保護に関する法律」(以下「個人情報保護法」)を改正する取り組みが進んでいる。日本政府は2014年6月までに大綱をまとめ、2015年に法改正をする予定だ。

パーソナルデータ利活用 のための 「匿名化」について議論

NIIの佐藤一郎教授は、この個人情報保護法の改正の大枠を作るために、2013年9月に発足した内閣官房「パーソナルデータに関する検討会」に参加し、さらに同検討会の「技術検討ワーキンググループ」の主査を務めて、パーソナルデータの「匿名化」などに関して検討した。ここでの検討結果は、「技術検討ワーキンググループ報告書」として2013年12月に公表されている。

佐藤教授は、CPSにおいてもプライバシー保護の問題は非常に重要だと話す。「CPSの技術が広まることで、取得したデータから個人の特定がやりやすくなり、プライバシー問題を引き起こす可能性がある。そこで、『技術検討ワーキンググループ』では、最新の技術を前提に対策を議論しました」と、佐藤教授は言う。

CPSが本格的に利用されると、センサーが現在よりも広範囲に使われるようになる。データを収集して利用する事業者も増える。インターネットをはじめとする、膨大な他のデータとの突き合わせも考慮すべきだろう。このように最新の技術を採用するうえで弊害が出てくれば、それを解決しなければならない。CPSそのものに関する研究と同様に、こうした取り組みは、技術に携わる者として重要なことだと、佐



図：提案中の新たな個人情報利用のルール。匿名化を施したパーソナルデータである「(仮称) 法第23条1項適用除外情報」の第三者提供を例外的に認める。

藤教授は言う。

そうした背景のもと、先述の「報告書」ではプライバシー情報を守るという見地から、パーソナルデータの範囲の明確化、匿名化、第三者提供について技術的な検討を行っている。

CPSやビッグデータ技術が進むなかで、1つひとつの情報ではプライバシー上の問題が生じていないデータでも、センサーから取得した現実世界の情報やインターネット上にある多様な情報との突き合わせにより、個人を特定できる可能性が出てくる。そこで報告書では、今後、パーソナルデータとして考える範囲を、個人が誰であるかがわかるという「特定性」だけでなく、誰だかわからないが、ある情報が誰か1人に関することだとわかる「識別性」についても、明示的に扱うことが必要だと提言している。

例えば、誰とはわからない購買履歴でも、マイナーな商品を買っていると、買った人のブログなどと付き合わせれば、その商品を買った人が誰か特定できることもあるし、その人が他に買っている商品もわかってしまう。実際に、ネット通販業者などが、利用者の氏名から、本人のTwitterやFacebookを探し、そこから住所や趣味、生活行動を推定することは珍しくない。

これに対して、われわれの大切なパーソナルデータを守る方法として注目されているのが、匿名化だ。匿名化とは、パーソナルデータを加工することで、個人の特定性や識別性を減らす技術のこと。ただし、「匿名化に過度な期待は禁物」と佐藤教授は指摘する。「報告書」では、「個人が識別される（又は個人が特定される）情報が一切推定されない合理的な匿名化水準を汎用的に達成可能にすることは不可能である」と明記している。

一方で、パーソナルデータの保護と利活用はトレードオフとなる。これは特定性や識別性を抑えたパーソナルデータにおいては、利活用の価値も下がるということの意味する。前述の購

買履歴でも識別性を下げるには、購入者の氏名だけでなく、個々の商品を特定する情報、例えば商品の型番などを隠すなどの措置が必要であり、情報としての価値が下がっていくことになる。

ケースバイケースの対応が必要

もっとも、パーソナルデータを匿名化して、個人の特定性や識別性を減らせば、第三者提供を含む、パーソナルデータの利活用に道を開くことになる。パーソナルデータの利活用を推進する立場の事業者は、パーソナルデータの「匿名化のガイドライン」が明文化されることで、それを適用して事業者間でデータを交換して分析を進めるなど、利活用が進むことを期待していたのかもしれない。だが、技術的な見地から検討した「報告書」では、ガイドラインなどで一律に問題を解決しようとするのではなく、「ケースバイケースでの対応が必要」と指摘している。

さらに、「1人ひとりが識別できず、したがって特定もできない」ように匿名化するのは容易ではない。「報告書」では、Suica乗車券のような交通系ICカードの乗車履歴に関する考察を例に挙げている。その識別性を下げるための措置として、個人に結びつく「ID」を削除したり、乗降数が少ない駅のデータを除外したりする措置を施したとする。これで一見、良さそうだが、乗降履歴というのは乗車駅と降車駅の組合せからなり、乗降者数の多い駅から駅への移動でも、その駅の組合せに該当する利用者が多いとは限らない。仮に1人しかいない場合、識別できる状態となり、前述の措置は十分とはいえない。さらに乗降客数は日々変わるので、匿名化の仕方も変える必要がある。「残念だが、すべての事業者が適切に匿名化を行えるとは限らない。それを考慮した制度設計が必要」と佐藤教

授は強調する。

パーソナルデータの利活用はできるのか

当然、パーソナルデータの第三者提供では「特定性」や「識別性」をなくすことが前提となるが、匿名化が難しければ、第三者提供は困難となり、利活用が進まないことになる。そこで、報告書では第三者提供に関わる、パーソナルデータの新しい類型を定義している。これは氏名の削除など、ある程度の匿名化を行ったパーソナルデータについては、特定の第三者に対して提供を例外的に認める一方で、提供を受けた事業者は、誰に関する情報なのかを特定しないことを規律で制約する考え方だ。

また、個人情報保護法の改正では、パーソナルデータのコミッショナー役として、第三者機関の設置が議論されている。法に沿ってパーソナルデータの利活用を監視し、違反があれば立ち入り調査を含む必要な措置をして、それを正す。さらには、匿名化に携わる技術者集団が必要になるという。

「『攻める』側と『守る』側の立場を考えると、多くの場合、『守る』側の技術の方が難しいのです。セキュリティ問題が典型的な例ですが、プライバシー保護でも同じことが言えます。データサイエンティストの需要が高まるなら、それと同じように匿名化を支える技術、匿名化を行う技術者も求められます」と、佐藤教授は言う。

パーソナルデータの利活用とプライバシーの保護を両立させる難しさに対して、ケースバイケースで対応することが、技術者にとって今後の大きな課題となるのだろう。つまり、パーソナルデータが、技術者にとっての新たなフロンティアになるかもしれないのだ。

(取材・文＝星 暁雄)

圧縮・復元技術の最前線

センシングデータのスマート化に不可欠な技術とは

センシングデバイスから収集される膨大なデータを効率的に蓄積、かつ、高速で処理するためには、圧縮・復元技術が不可欠となる。だが、従来の圧縮・復元技術は、その処理を行うに際して、さまざまな課題を有していた。そうした課題を解決するため、より効率的な圧縮・復元技術の実現に向け、NIIの定兼邦彦准教授と北海道大学大学院の喜田拓也准教授らによる共同研究が進められている。

センシングデータの 効率的な 圧縮・復元が課題に

近年、ビッグデータの中でも、GPS（全地球測位網）やカメラ、携帯電話、スマートフォンなどのデバイスから得られる「センシングデータ」に注目が集まっている。CPS（サイバー・フィジカル・システム）の実現には、それらの多様なデバイスから連続的に収集される膨大なデータを効率的に蓄積するとともに、自由に取

り出してコンピューター上で活用できる仕組みが不可欠だ。そして、そのためにも、より効果的な圧縮・復元技術が求められている。

しかし、大量のデータを効率的に処理したり、高速に検索したりするためには、従来の圧縮・復元技術には限界があった。NIIの定兼邦彦准教授は、「大量のデータを高速処理するにあたり、これまでの圧縮技術では、圧縮データをコンピューターのメモリ上でそのまま読み書きできないという課題がありました。また、特定のデータへのランダムアクセスについても、非圧縮データであれば容易に行えますが、圧縮データでは不可能なため、先頭から1つずつ復元しなければなりません。対して復元の時間を短くするには、データを複数のブロックに分割し、個々に圧縮しなければならず、結果、圧縮率の低下を招いていました」と説明する。

北海道大学大学院の喜田拓也准教授も、「データ量が少なければ、圧縮や復元を行うことなくそのままメモリ上で処理できます。しかし、センシングデータのような膨大な量の情報を圧縮・復元せずに高速に処理するためには、大容量のメモリを搭載した大規模なコンピューター環境が必要となります。また、データを保管しておくためのハードディスクの容量もおのずと増大してしまいます。そうしたことから、従来はある一定の近似値でデータをまとめた

り、平均値だけを保存したりするといった手法がとられていましたが、その過程で間引かれたデータに存在する知見を見逃してしまうということになってしまいます」と補足する。

「このような課題の解決に向け、喜田先生、九州大学の竹田先生とともに効率的な圧縮と復元を可能とする基盤技術の研究を、2012年より共同で進めてきました」（定兼准教授）

VF符号と Re-pairアルゴリズム による圧縮を推進

では、効率的な圧縮・復元の実現に向け、現在、どのような共同研究が進められているのか。喜田准教授が手掛けているのが、「VF符号（Variable-length-to-Fixed-length code）」と「Re-pairアルゴリズム」を用いた圧縮・復元である。VF符号とは、元のデータにおける長さの異なる部分系列に対して、固定長の符号を割り当てる符号化方式だ。

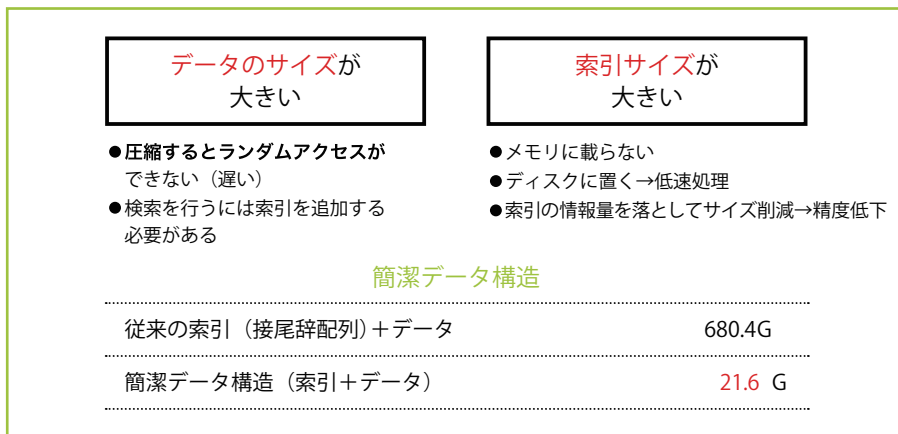
「例えば、あるデータの1つの塊が5bitで表現されている場合、圧縮されているデータの中から5番目の符号語で表現されている情報だけを抽出したい場合、21番目～25番目を取り出



喜田拓也 Takuya Kida

北海道大学 大学院情報科学研究科
コンピュータサイエンス専攻 准教授

図1: 従来のデータ構造の問題点





Replace most frequent bigram with a new symbol

AA ABAC AA ABCC AA AB
 ↓
 DA BAC DA BCC DA B
 ↓
 EB AC EB CC EB
 ↓
 FA CF C CF
 ↓
 FAGCG

dictionary

D → AA
 E → DA
 F → EB
 G → CF

compressed text

図2: Re-pair アルゴリズムの仕組み

せば済むわけです。これが可変長の符号語で圧縮されたデータでは、前から順番に見ていかなければなりません」(喜田准教授)

このように、符号語の境界が明確であることから、圧縮されたデータへのアクセスが容易というメリットをもつが、VF符号は可変長符号に比べて圧縮率を向上させることが難しい、という側面があった。喜田准教授は「そこで、Re-pair アルゴリズムを用いた文法変換に基づく圧縮方法を組み合わせることで、圧縮の効率を保ちながら、データを復元することなしにキーワード検索等の処理を高速に行える仕組みを実現しています(図2)」と説明する。この組み合わせにより、一般的な圧縮ツールである「gzip」を上回る圧縮率が実現されている(図3)。

簡潔データ構造 による 新しい圧縮・復元法

一方、定兼准教授は、「簡潔データ構造」に基づいた圧縮・復元技術の研究を推進している。「私は、簡潔データ構造と呼ばれる新しい圧縮法を研究していますが、その簡潔データ構造において、文字列を圧縮したまま高速に検索で

きる圧縮接尾辞配列を開発しました」と、定兼准教授は話す。

「文字列の検索用の索引をコンピューター上へ実装しようとした場合、余分なデータを付加してしまうことが多く、結果、索引が元データよりも大きくなってしまいがちです。この問題を解決するために圧縮接尾辞配列が提案されていますが、検索にはテキスト自体が必要となるため、索引のサイズがテキストよりも小さくならないという課題がありました。そこで、私は圧縮接尾辞配列を用いた検索アルゴリズムにおいてテキスト自体が不要になるように変更を加えたり、テキスト全体やその一部を圧縮接尾辞配列から復元するアルゴリズムを用いることを提案しました。これにより、テキストや索引の圧縮に加え、任意の語句の検索や文書の任意箇所の部分復元を実現しました」(定兼准教授)

喜田准教授は、「定兼先生が研究している簡潔データ構造を活用した圧縮・復元技術を用いることで、メモリ上に乗せられるほどデータをコンパクト化できるため、データアクセスが容易になります。すなわち、大容量メモリを搭載した大規模なコンピューティング環境を用意せずとも、膨大なデータの処理や分析が可能となるわけです」と強調する。

現在、テキストデータを対象とした圧縮・復元の研究が共同で進められており、期待される実用例の1つとして、DNAやヒトゲノムのデータ処理などが挙げられている。また、今後はセンサーデータへの適用も進めていき、まずは、車載位置情報システムから収集されるデータの圧縮・復元に着手していくことを予定しているという。

新しい圧縮・復元技術により、センシングデータの処理がより高速に行えるようになれば、膨大なデータの中に埋もれ、これまで見過ごされてきたようなさまざまな知見が得られるようになるだろう。その実現に期待が寄せられる。

(取材・文=伊藤秀樹)

(よく知られた gzip という圧縮ツールよりも圧縮率が良い)



図3: Re-Pair アルゴリズムとVF符号化の組み合わせにより、圧縮率が向上



定兼邦彦 Kunihiko Sadakane

国立情報学研究所 情報学プリンシプル研究系 准教授
 総合研究大学院大学 複合科学研究科 情報学専攻 准教授

都市社会のOS

田中 謙 Yuzuru Tanaka

北海道大学 大学院情報科学研究科 教授



人口100万人以上の都市の中で、年間降雪量が世界一はどこかで存知だろうか？ 札幌である。札幌の人口は190万人、年間降雪量は6m。一晩の車道除雪の距離は札幌と石垣島の往復距離に達す。札幌市は年間150億円を除排雪に支出。昨年度は220億円に膨らんだ。ICTを用いた最適化でコスト削減を図るのは自然である。20年以上前に試みがあったが望ましい成果は得られていない。

一方、この20年間にICTは革新的に変わった。ウェブの発展、GPS付きスマホやカーナビの普及、位置情報活用技術の進展、渋滞センサーの普及、大規模データの配信・蓄積・管理検索・分析技術の高度化等である。

カーナビやスマホ、渋滞センサー等、都市に遍在するセンサー群と気象レーダなどを活用して物理世界をモニタリングできる。除排雪記録等のデータベースも利用できる。道路の積雪・堆雪状況は車載レーザ・レンジ・スキャナで実測できる。気象変化と除排雪履歴の路面凍結や路幅狭隘への影響を、プローブ・カー・データから実時間推定し、除排雪の緊急度を定量的に評価することは可能であろう。交通事故との相関分析により、危険な道路交通状況を刻々とドライバーに伝えるのも難しくはない。

物理世界のモニタリングによりサイバー空間内にモデリングを行い、関連DB（データベース）と共に用いて、対象世界の最適制

御を目指すシステムをサイバー・フィジカル・システム（CPS）という。対象世界が心臓のペースメーカーも、大規模プラント制御システムもCPSである。ここでの関心は、対象世界が都市社会に拡がったソーシャルCPSである。

類似の構想は、古くは1991年にD・ガランターが著した『ミラーワールド』に見られる。ミラーワールドとは、物理世界を鏡のように忠実に映したサイバー空間のこと。両世界の整合性を取るインタフェースとして考案されたのが「ダブル・スペース」（共有メモリ空間）だった。

ソーシャルCPSはビッグデータ・システムというより都市社会のOSだ。都市社会のリソース・スケジューラであり、モデリングでリソース仮想化を進めて分析制御システムの開発・適用を容易にする。意思決定への人の介在を支援するユーザ環境も提供する。

持続可能で安心安全な都市社会の構築に、ソーシャルCPSの必要性は増している。必須要素技術も急速に成熟している。残るは、民間や行政がサイロの中で保持するデータのオープン化と、本質的にシステム・オブ・システムで、単一モデルでの全体記述が不可能な大規模複雑系のデータ分析をどう行うかである。ソーシャルCPSは魅力的な研究開発課題に溢れているように思う。

情報から知を紡ぎだす。

表紙イラスト

実社会とサイバー空間がリアルタイムでつながると、今、どこで何が起きているのかが、一覧できるようになる。「犬を連れて富士山周辺へ行きたい」というニーズに対して、犬と遊べる施設や道路の状況、その他スポット情報などが手に取るようになる。そんなCPSの実現は近い。

国立情報学研究所ニュース [NII Today] 第63号 平成26年3月

発行：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 <http://www.nii.ac.jp/>

〒101-8430 東京都千代田区一ツ橋2丁目1番2号 学術総合センター

編集長：東倉洋一 表紙画：小森誠 写真撮影：川本聖哉 / 佐藤祐介 デスク：田井中麻都佳 制作：クディラアンドアソシエイツ株式会社
本誌についてのお問い合わせ：総務部企画課 広報チーム TEL：03-4212-2164 FAX：03-4212-2150 e-mail：kouhou@nii.ac.jp