実世界情報の可視化分析基盤

ーセンサーデータから見て解く環境と社会のつながりー

村里 英樹（むらさと ひでき）
ユニバーサルコミュニケーション研究所 情報利活用基盤研究室 有期技術員

大学院修士課程修了後、国際電気通信基礎技術研究所（ATR）研究技術員を経て、2009年より現職。
情報可視化システム、検索インタフェースなどに関する開発に従事。

背景

近年、自然現象や社会現象など実世界の状況を反映したデータ（ここでは以下「センサーデータ」という）が急速に拡大しています。オープンデータ活動の進展や参加型センシング、ソーシャルセンシング技術などを通じて、ユーザが利用可能なセンサーデータは今後も増え続けていくでしょう。

NICTではこのようなセンサーデータを利用し、実世界の様々な出来事から情報を収集・探索・発見・共有するための参加型センシング基盤を開発しています。本基盤のシステムはデータの収集、蓄積、検索、可視化モジュールで構成されます。本稿では情報探索、情報可視化の観点からご紹介します。

異分野センサーデータのVisual analytics

Visual analyticsは、大規模かつ複雑なデータを利用して効率的な情報獲得や意思決定を行う分析プロセスです。自動化された統計学やデータマイニングの分析手法と、対話的な視覚的インタフェースを相補的に使用して、どちらか一方だけでは難しい課題の解決を図ります（図1）。

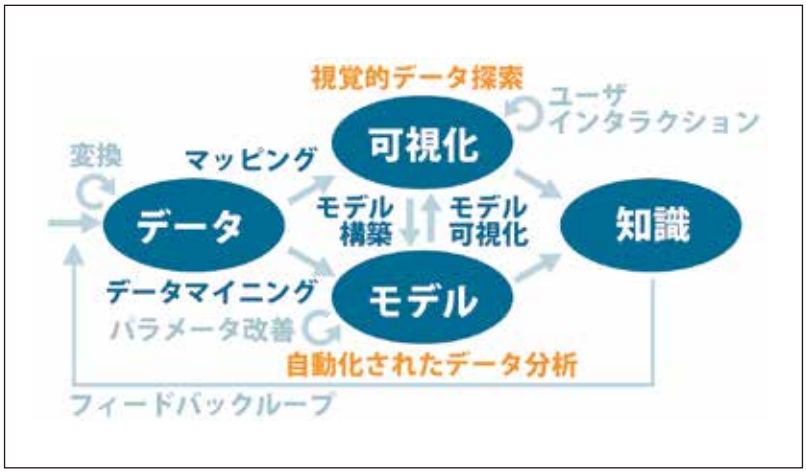

図1　Visual analyticsのプロセス
知識発見のためのデータ、可視化、モデル、ユーザの相互作用に特徴がある。

異分野のセンサーデータから相関が見られるセンサーデータの組み合わせを発見し、また相関が見られる要素（時間、空間、値域、キーワードなどのパラメータ）の特定を行う相関分析を実現するため、センサーデータ検索技術とともに、データの時空間パターンを対話的に可視化する時空間情報可視化システムSTICKER（Spatio-Temporal Information Clustering and Knowledge ExtRaction）を開発しています。これにより、科学的モデルを作ることが困難な異種・異分野データの相関関係を視覚的に分析することが可能となります。

STICKERの特徴

STICKERでは地理空間上の分布の時間に沿った変化を確認できるようにするため、地理空間と時間軸から構成される3次元空間にデータを表示します。この3軸での表示はSpace time cubeの名前で知られる一般的な方法ですが、STICKERには異なる分野のセンサーデータの情報探索を目的とする以下のような特徴があります。

さまざまな分野のデータに対して横断的な操作や可視化を可能にすべく、STT（空間、時間、テーマ）形式（図2）でデータを統一的にモデル化し、各種データをSTT形式に変換して様々な形状の可視化オブジェクトの生成とそれらの視覚操作、STT属性に基づくデータの選択や集約などの各種のデータ操作を実現しています。これらのデータ操作の結果を即時に表示することで、特徴的なパターンが強調されるフィルタパラメータやその組み合わせの発見を効率的に行うことができます。

また時間空間を各軸に沿って等間隔に区切り、ブロック毎にデータの集約結果を保持するSTTセル形式によってデータの粒度をコントロールし、STICKERと同じくSTT形式でセンサーデータを蓄積するEvent Warehouseと連動したスケーラブルなデータ処理、表示の仕組みを実現しています。表示方法多様化にもこのSTTセル形式は有効であり、

異なる表示方法を混在させることで、複雑になりがちなセンサーデータの種類の識別を助けてセンサーデータの時空間パターンの把握を容易にしています。

図2　STT形式に基づくSTICKERのVisual analyticsデータ処理
STT形式によって横断的なデータ操作と可視化を可能にし、操作を表示に動的に反映する仕組みでパラメータ修正の繰り返しを効率化。

探索結果の例

本システムで探索した例を2つご紹介します。

図3は「日本でPM2.5と相関が強い現象と時期」について探索した結果です。2013年8月の日本における大気品質（PM2.5）および降水の観測データと、渋滞に関するキーワードを含むTweetの分布を表示しています。赤色で示すPM2.5が35μg/m³以上の領域は8月初旬には全国的に広がるものの中旬で途絶えており、水色で示す降水量2mm/h以上の領域と補いあうように変化していることが確認できます。また灰色で示す渋滞のTweetが存在した領域は中旬頃に東日本から西日本にかけて広く分布しており、お盆休みのUターンラッシュのピークと重なります。例えば以上の観察から「夏季休暇に伴う交通集中はPM2.5を35μg/m³まで増加させた」といった仮説を立てることができ、これをひとつの情報として探索を継続する、補強する例や反例を別の年や場所や現象から探すといった、より精緻な分析のための手がかりを見つけ、その分析で使用するデータ・セットを取得することができます。

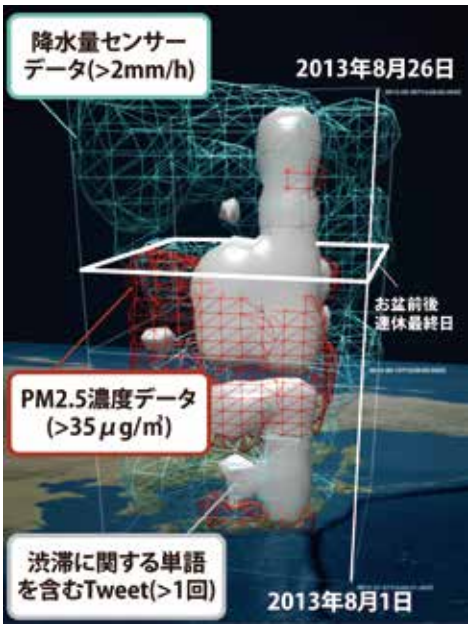

図3　2013年8月の日本における現象を表すデータ
赤色の網はPM2.5が35μg/m³以上、水色の網は降水量が2mm/h以上、灰色の面は渋滞に関するキーワードを含むTweetがSTTセル当たり1回以上の領域。

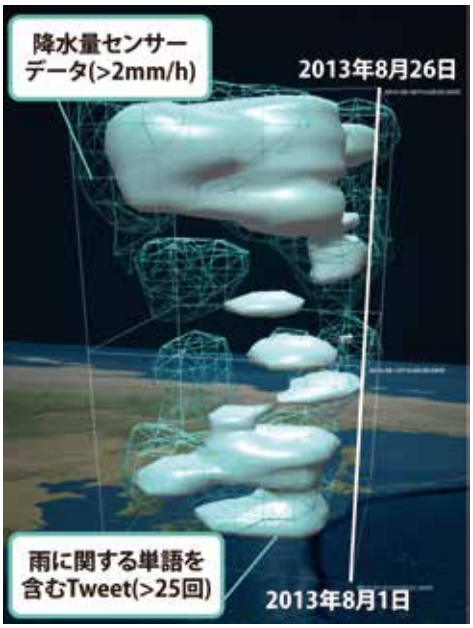

図4　2013年8月の日本における現象を表すデータ
水色の網は降水量が2mm/h以上、水色の面は雨に関するキーワードを含むTweetがSTTセル当たり25回以上の領域。

図4は「日本で相関が強い自然現象とソーシャルメディアデータの組み合わせおよびその時期」について探索した結果です。2013年8月の日本における降水量と雨のキーワードを含むTweetの分布を示しています。2つはこの時間空間の全域で重なり合っています。このように、互いに代替可能なセンサーとして利用の幅を広げられる可能性があるセンサーデータの組み合わせやその値域を一目瞭然に確認することができます。また雨に関するキーワードを含むTweetに対し、雨の降り方が激しい領域とそうでない領域の間でTweet内容の差異を比較することで、雨の降り方に対する人々の反応の違いを見ることができます（例えば、どの地域で何mm/h以上の雨が降ると「ゲリラ豪雨」に関するメッセージが増えるか、など）。

今後の展望

参加型センシング基盤は、様々なセンサーデータを皆で協調して収集・分析することを目的にNICTの知識・言語グリッド上で開発が進められているシステムで、ユーザ自身が様々な情報源からデータを収集するための情報サービスを開発するユーザ定義センサー、STT形式に基づいてこれらのデータを蓄積・統合するEvent Warehouse、センサーデータ検索技術、および時空間情報可視化システムSTICKERによって構成されます。今後は、この参加型センシング基盤を使って、自然現象から社会現象まで様々なセンサーデータを網羅的に収集し横断的に分析することで、大気汚染や気候変動などの環境問題と社会への影響をユーザが協力して多角的に分析する情報基盤の実現に取り組んでいきます。

また、より多くのデータを対象に可視化分析を高速かつ効率的に行えるようにすべく、時間・空間・概念的に相関の強いデータを検索する機能（相関検索機能）や、膨大なセンサーデータの中からあるイベントの発生を示すデータのみを選択して高速に可視化処理する機能（異常値検出機能）の研究開発を進めています。さらに、可視化分析の途中で不足しているデータに気づいたらその場でユーザ定義センサーを作成して集めたり、複数のユーザ間で分析結果を共有しながら協調して分析ができるようにするなど、可視化分析のユーザビリティ向上にも取り組んでいく予定です。