3.5.4　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

研究室長　　是津耕司　ほか 10 名

センサーデータからソーシャルデータまで網羅した異分野ビッグデータの横断的利活用基盤

【概　要】

インターネット上でアクセス可能な膨大なテキスト、マルチメディア、センサーデータなどの情報コンテンツや、情報コンテンツの一種と見なすことができる情報サービスを組み合わせ、ユーザの要求に対して、広い観点に立った、効率の良い意思決定を支援する情報利活用基盤を開発する。具体的には以下の研究開発を行う。

(1)　大量かつ多様なテキストやセンシングデータから構築された大規模情報資産の管理技術を開発する。

(2)　大規模情報資産を利用する情報サービスの検索や管理を行い、適切な連携をすることでユーザの要求を満たす複数のサービスを発見し、それらのサービスを適切に組み合わせて効果的に実行させる情報サービス連携技術を開発する。

これらの技術に基づき、センサーデータからソーシャルデータまで、異種・異分野のデータを横断的に検索・統合・可視化するシステムを開発する。また、これらを用いて情報利活用サービスを開発するためのプラットフォーム（知識・言語グリッド）を JGN-X 上に開発する。

【平成 26 年度の成果】

センサーデータや科学データ、ソーシャルデータなど 150 種類・180 万件のオープンデータを登録した情報資産リポジトリを構築するとともに、環境問題分野への応用に向け、これまでに開発した基盤技術の性能改善や応用システムの開発に取り組んだ。分野横断相関検索システム Cross-DB Search では、ある現象の観測データの周辺から、その現象の原因や影響になりそうな他のデータを芋づる式に発見できるようにすべく、時空間・テキスト疑似適合性フィードバック手法 STT-PRF を開発した。STT-PRF では、キーワード検索結果上位のデータ（シードデータ）から、それらが作成された場所や日時の情報をフィードバックすることで元のクエリを拡張し、キーワードにヒットするデータだけでなく、その周辺データまでも自動的に検索する。その結果、例えば“森林破壊”というクエリに対し、南米地域では乾燥化、アフリカ地域では土壌劣化や砂漠化、東南アジア地域では汚染など、地域ごとに異なる相関データを発見することが可能になる。こうした検索は、環境問題など、様々な分野にまたがり、かつそれらの関係性が十分に解明されていないような課題を扱う際に、実世界で空間・時間・テーマ的（Spatial, Temporal Thematic: STT）に相関が高いデータを網羅性高く発見し、分野横断的な分析の手掛かりを見つけることに役立つ。評価実験では、気候変動や大気汚染など環境問題に関するクエリ 50 個に対し、従来のキーワード検索と比べ、Normalized Discounted Cumulative Gain 比（nDCG@30）で平均約 6 % の検索精度改善を達成した。こうした性能改善を施した Cross-DB Search を使って、国内最大級の科学技術データ事業者である科学技術振興機構（JST）との共同研究に基づき、約 3,000 万件の科学技術データを対象とした相関検索システムを開発し、研究動向調査などへの応用に向けた内部検証に着手した。

一方、様々な分野にまたがるセンシングデータを横断的に統合し、実世界における環境変化や社会の動きの相関を網羅的に分析する Event Data Warehouse 基盤の開発にも取り組んだ。複合イベント解析技術は、あるセンシングデータと時間・空間・テーマ的に相関が高い他のセンシングデータを発見し、ある事象（イベント）に関連する様々なイベント（複合イベント）の発見に役立てる。物理センシングからソーシャルセンシングまで、種類の異なる多種多様なセンシングデータの相関を横断的に分析できるようにすべく、センシングデータの値の変化パターンを SAX 法（Symbolic Aggregate approXimation）で共通の形式に記号化し、異種センシングデータ間の横断的な相関ルール抽出を可能にする方法を開発した。これにより、例えば、降雨データの値が急に変化した場所・時間の周辺で、同様に出現頻度が急増した様々なトピックのソーシャルデータ（洪水や渋滞、雨宿りなどに関する SNS データなど）をまとめて発見することができ、ゲリラ豪雨がもたらす様々な影響を網羅的かつ精度高く把握することが可能になる（図 1）。ソーシャルデータのみから抽出する場合に比べ、物理センシングデータとの相関を取ることでノイズを削減し、抽出精度を約 30 % 向上させた。また、欠損データを補完したり未知データを予測し複合イベント解析の性能を向上させるべく、物理センシングからソーシャルセンシングまで、複数のセンシングデータの相関性を DRNN（動的リカレントニューラルネットワーク）に基づく深層学習によりモデル化し未知データを予測する方式を開発した。大気汚染を示す指標である PM2.5

を対象とした評価実験では、過去の PM2.5 データと気象データ（気温、湿度、風速、降雨など）の相関を学習し、従来の科学モデルに基づく大気汚染予測システム（VENUS）を上回る性能を出せることを確認した（過去 48 時間分のデータを学習）。

情報サービス連携技術の研究開発においては、これまでに開発した知識・言語グリッドをモバイル・ワイヤレステストベッド（**3.13** 参照）上に展開し、異分野センシングデータの収集解析を行うための応用システムを開発した。この応用システムでは、Service-Controlled Networking（SCN）技術による網内データ処理と自動パス選択により、無駄なデータ転送を抑えネットワークの輻輳を回避することができ、評価実験では、従来のベストエフォート方式に比べ、より多くのデータ収集処理をより長い時間、安定的に稼働させられることを確認した。また、データアクセスプロトコル処理、データ構造解析、データ保存処理など、データ収集に共通する処理をあらかじめ実装したソフトウェア開発キットをオープンソースとして公開し、ユーザ独自のデータ収集解析アプリケーションを容易に開発できるようにした。さらに、これまでに開発した分野横断相関検索システム Cross-DB Search と統合可視化分析システム STICKER の Web サービスを公開し、アプリケーション開発者がこれらの Web サービスを独自のアプリケーションに組み込み（マッシュアップ）、収集したデータの検索や可視化を容易に行えるようにした。これらにより、今中期計画最終年度である平成 27 年度の応用実証に向けた、アプリケーション開発プラットフォームの準備が整った。

さらに、NICT 内外との連携プロジェクトや研究協力を通じた成果展開を積極的に行った。ソーシャル ICT 推進研究センターと連携し、ゲリラ豪雨被害パトロール支援を目的とした豪雨早期探知情報配信と市民からの被害情報の収集及び統合可視化分析の実証システムの開発に着手した。一方、世界科学データシステム WDS との連携や JST との共同研究に基づく科学技術データ利活用への取組が、首相官邸 知的財産戦略本部が主催する知的財産推進計画 2014 の施策項目「アーカイブの利活用促進に向けた整備の加速化」の項目の 1 つとして登録され、定期的に成果報告を行った。また、国際科学会議 ICSU の科学技術データ委員会 CODATA 及び標準化タスクグループを通じ、WDS 連携に基づくデータサイテーション分析の研究成果を発表したところ、「データサイテーションを進める上で日本から重要な発表がなされた。皆で拍手を送りたい。」と異例のコメントがあり、日本学術会議にも報告されるなど大きな反響を得た。さらに、米国標準技術院（NIST）との共同研究では、Cyber-Physical Cloud Computing（CPCC）に関する国際ワークショップを大阪で開催するとともに、日米インターネットエコノミー政策協力対話でこれまでの NIST との共同研究の成果を発表し、日米政府間で CPCC の共同研究開発協力への合意を実現した。その他、国際標準化に向け、ITU-T スマート・サスティナブル・シティ標準化（FG-SSC）における異種・異分野の情報コンテンツの構築・統合・検索・配信のためのインフラ技術の策定に寄与した。

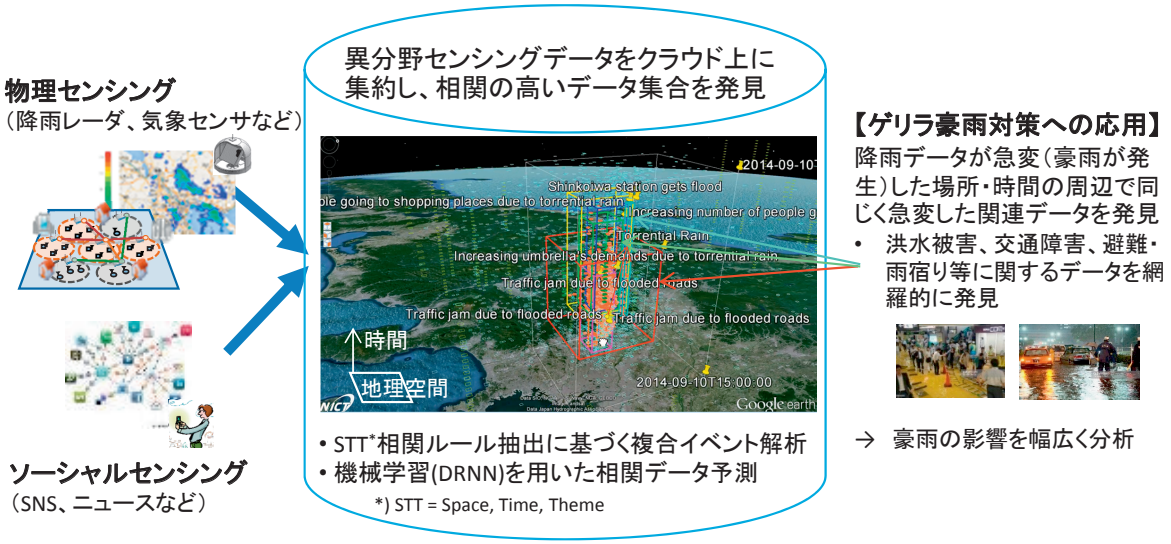

図 1　異分野センシングデータ統合解析基盤 Event Data Warehouse