

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

室長 是津耕司 ほか 13 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る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 25 年度の成果】

【情報資産管理技術の研究開発】

センサーデータやソーシャルデータなど実世界を反映したデータを対象に、異分野データの時空間相関を可視化しながら相関データをインタラクティブに発見する可視化分析技術を開発した(図 1)。これまでに開発した時空間相関可視化技術 STICKER を拡張し、異種・異分野の実世界データを STT スキーマ (Space、Time、Theme) に基づいて横断的に検索したり可視化したりデータ操作(集約、フィルタリング等)を行ったりする方法を提案し、これを実装したイベントウェアハウス基盤を開発した。これを用いて、様々な分野のデータの中から相関がありそうなデータの組み合わせを見つけ出し、かつ相関が見られる時空間範囲や各データの閾値等

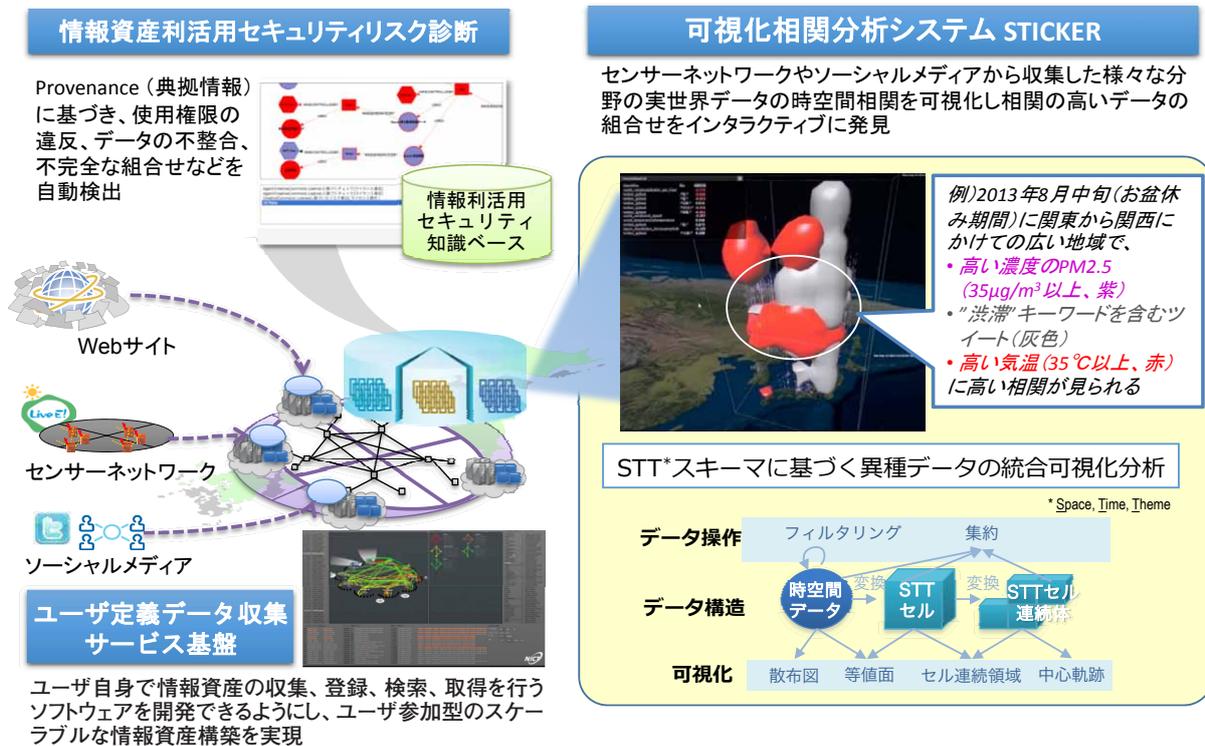


図 1 分野横断的な相関データをインタラクティブに発見する可視化分析技術

を調整しながら相関データを絞り込む可視化相関分析 (visual correlation analysis) 手法を提案し、STICKER に実装した。特に、センサーデータからソーシャルデータまで様々な種類のデータの間の時空間相関を横断的に分析するため、散布図や等値面、セル連続領域、中心軌跡など多種多様な手法で各種データの時空間的な広がりや連続性、動きを可視化し、それらの重なりや同期をインタラクティブな操作で視覚的に把握できるようにした。このようなセンサーデータからソーシャルデータまで幅広いデータを統合し時空間相関を可視化分析する技術は他に類を見ず、その特徴を生かし大気汚染と環境や社会との横断的な相関を調べる用途に応用し、例えば $35\mu\text{g}/\text{m}^3$ 以上の高い濃度を示す PM2.5 データと渋滞に関するキーワードを含む Twitter データ、および 35°C 以上の高い気温を示す気象データを、平成 25 年 8 月中旬のお盆休み期間中に関東から関西にかけての広い地域で高い相関を示す組み合わせとして発見するといったようなことが可能になった (図 1)。

その他にも、どの情報サービスがどの情報資産を利用しているのかや、ある情報資産が他のどの情報資産を組み合わせで作られているのかなど、情報資産利活用の provenance (典拠情報) を W3C Open Provenance Model 標準に基づいて構造化した上で、情報資産提供者・利用者間での使用権限の違反、データの不整合、不完全な組み合わせなど、情報資産利用に関する 7 種類・100 項目以上に渡るセキュリティルールを定義した知識ベースを構築し、情報サービスの開発時や実行時に provenance を解析して情報資産利用のセキュリティ違反を自動検出する技術のプロトタイプを開発した。これにより、情報資産利用におけるセキュリティリスクの“見える化”を可能にし、情報サービスの開発時や実行時にセキュリティリスクを動的に診断できるようにすることでユーザが情報資産をより提供しやすくなった。

【情報サービス連携技術の研究開発】

これまでに、NICT が運用する広域テストベッドネットワーク JGN-X 上に開発した知識・言語グリッドを拡張し、ユーザ所有のサーバー上で情報資産の収集、登録、検索、取得を行うユーザノード用ソフトウェアを開発し、ユーザノードが知識・言語グリッドネットワークに接続することでグリッドの規模を拡大し、より多くの情報資産を作成・共有できるユーザ参加型のスケーラブルなシステムを開発した。また、この上で、元データを継続的に収集し情報資産に登録するユーザ定義センサー (センサーデータやソーシャルメディアデータ等の情報資産を対象) や、複製できないデータのメタデータを抽出し登録するユーザ定義ハーバスタ (科学データ等の情報資産を対象) をユーザが独自に開発し実行できるようにするデータ収集サービス基盤を開発し、共通機能のライブラリ化等によってユーザによる収集プログラムの開発コストを約 9 割削減することに成功した。これらにより、75 種類・125 万データセット規模の情報資産を構築し、情報資産の作成効率を昨年度に比べ約 2.6 倍 (データセット数に基づく比較) に改善し、情報資産のスケーラビリティを大幅に向上させた。さらに、新世代ネットワーク連携プロジェクトで開発中のサービス要求に応じたネットワーク構成の自動設定技術 (Service-Controlled Networking ミドルウェア) をデータ収集サービス基盤に適用することで、指定されたスループット要求を保持しつつ同時実行可能なサービス数を増加させられることをシミュレーション実験等で確認し、情報資産のスケーラビリティ向上に対する有効性を示した。

【機構内外との共同研究】

NICT が国際プログラムオフィスを務める World Data System (WDS) との連携では、ICSU CODATA Task Group on Data Citation Standards and Practices に参画し、データサイテーションに基づく科学データの相関検索など研究成果の一部を盛り込んだ標準化報告書を出版した。また、科学技術振興機構 (JST) とビッグデータ利活用に関する MOU を締結し、相関検索エンジン Cross-DB Search を JST 科学技術データの検索に応用したシステムのプロトタイプ開発に着手した。平成 23 年度に MOU を締結した米国商務省標準技術院 (NIST) との連携では、Cyber-Physical Cloud Computing の概念設計をまとめた技術報告書を共同で作成し公開した。また、NIST 連携から発展したカリフォルニア大学アーバイン校との研究協力では、Cyber-Physical Cloud Computing の実現に向けたイベント情報管理基盤技術の共同研究を実施した。こうした活動の成果を、インターネットエコノミーに関する日米政策協力対話 (第 5 回局長級会合) や日米 ICT R&D フォーラム (The 10th Annual Colloquium on ICT R&D) などでも発表し、Cyber-Physical Cloud Computing 分野に係る研究開発を日米政府で協力して推進する声明が出された。さらに、センサーデータ収集解析基盤を新世代ネットワーク連携プロジェクトで開発するとともに、うめきた超臨場感実証実験システムの 10 面 3D タイルドディスプレイにも STICKER の 3D 版を開発・導入した。