

日本—アメリカ 国際共同研究「ビッグデータと災害」 平成 28 年度 年次報告書	
研究課題名（和文）	災害時応用のための効率的かつスケーラブルなビッグデータの収集・解析・処理
研究課題名（英文）	BDD: Efficient and Scalable Collection, Analytics and Processing of Big Data for Disaster Applications
日本側研究代表者氏名	原 隆浩
所属・役職	大阪大学 大学院情報科学研究科・教授
研究期間	平成 27 年 4 月 1 日～平成 30 年 3 月 31 日

1. 日本側の研究実施体制

ワークパッケージ①	複数の情報源からの高次元データのモニタリング機構（WP3）	
氏名	所属機関・部局・役職	役割
原 隆浩	大阪大学・大学院情報科学研究科・教授	問題定義、方法論立案、機構の設計・拡張
西尾 章治郎	大阪大学・総長	機構の実装、実証実験

ワークパッケージ②	マイクロブログの実社会センシング応用のための機構（WP5）	
氏名	所属機関・部局・役職	役割
原 隆浩	大阪大学・大学院情報科学研究科・教授	問題定義、方法論立案、機構の設計・拡張
西尾 章治郎	大阪大学・総長	機構の実装、実証実験

2. 日本側研究チームの研究目標及び計画概要

平成 28 年度は、日本側研究チームは WP3「複数の情報源からの高次元データのモニタリング機構」において、初年度に実装した機構の評価・拡張を行う。人工データおよび実データを用いた詳細な評価実験を実施し、実装した機構の有効性の検証および問題点の抽出を行う。その結果に応じて、実装した機構の修正・拡張を行う。

さらに、今年度から新たに WP5「マイクロブログの実社会センシング応用のための機構」を推進する。WP5 では、マイクロブログのメッセージなどを社会センサとして利用可能にする機構について、メッセージに位置情報を付与する技術や、メッセージから興味・感情を抽出する技術を設計・実装し、さらに評価実験による有効性の検証を行う。また、社会センサデータを共有する基盤技術についても研究開発を行う。

3. 日本側研究チームの実施概要

平成 28 年度は、日本側研究チームは WP3「複数の情報源からの高次元データのモニタリング機構」において、まず、平成 27 年度に実装した検索機構のうち、詳細な評価を未実施だったものについて、大規模な人工データといくつかの実データを用いた評価実験を行った。その結果、多くの考案機構は、従来技術と比較して数 10 倍から 100 倍程度の性能を達成した。また、改善点が見つかった考案機構については機構の修正・拡張を行った。

さらに WP3 では、平成 27 年度に実装した標準的な検索以外に、既存研究では考えられていなかった新しい種類の検索技術について研究開発を行った。例えば、(1)検索結果となるデータ（位置依存データ）の地理的多様性を考慮した Top-k 検索、(2)MaxRS（固定サイズの領域で包含可能な利得最大となる場所を発見）の連続問合せ（モニタリング）について、有効な技術を考案した。これらの考案技術を実装し、性能評価を行った結果、単純な従来技術と比較して、数 10 倍から数 100 倍ほど高速に検索を実行できることを確認した。

さらに今年度から、新たに WP5「マイクロブログの実社会センシング応用のための機構」の研究開発を推進した。まず、マイクロブログデータを一種の社会センサとして扱うために、マイクロブログデータに位置情報（投稿位置）を付与する技術、および、マイクロブログデータからユーザの興味（話題）や世間のトレンドを抽出する機構を考案した。これらの機構では、特にマイクロブログデータが短文で文法が不正確な場合が多い、つまり、ノイズが多く情報量の少ない文書であることを考慮して、Wikipedia などの外部情報源を併用して語彙拡張やノイズ除去を行っている。

上記の研究において、マイクロブログデータのようなノイズの多い短文では、複合語の重要度を正確に測ることが困難であるという問題が頻繁に生じた。この問題を解決するために、当初の WP5 の研究計画には含まれていなかったが、複合語の重要性を測る新たな指標について研究を推進した。具体的には、情報距離の概念を導入し、（複合語ではない）単語の重要度指標として最もよく用いられる IDF を複合語にも適用可能なように拡張した指標を考案した。

さらに WP5 では、マイクロブログデータの解析結果（社会センサと呼ぶ）および解析手法の共有を可能とする社会センサデータ共有基盤システムの研究開発を進めた。これまでに、解析手法の登録や登録された手法の実行、解析手法や解析結果のダウンロードなどの最小限

の機能を実装したプロトタイプシステムの開発を完了した。さらに、ユーザインタフェースやデータフロー制御の機構について研究開発を進めた。

これらの研究成果は、国際論文誌 3 編、国内論文誌 3 編、国際会議 9 編、国内会議（査読有）5 編で公表もしくは公表予定であり、予定以上の大きな成果を達成した。これらの中には、世界最高峰の論文誌 2 編（ACM Transactions on Information Systems および ACM Transactions on Spatial Algorithms and Systems）、世界最難関の国際会議 1 編（Int. Conf. on Distributed Computing Systems）や、著名論文誌（World Wide Web）、著名国際会議（Big Data や DEBS）が含まれている。さらに、国際会議の最優秀論文賞を含めて 2 件の研究賞を受賞している。