

戦略的創造研究推進事業 さきがけネットワーク

研究課題「人流ビッグデータによる地球規模の
課題解決のための情報学と社会科学の融合基盤構築」

研究終了報告書

研究期間 平成29年12月～平成31年3月

研究代表者：水野貴之

(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 情報社会相関研究系、准教授)

§ 1 実施概要

(1) 実施概要

解析技術を開発する杉山グループと、解析結果の意味を言語化する技術を開発する宮尾グループ、それらの技術を使って社会問題を解決する水野グループが協力し、(a) 人流に関する汎用的なデータマイニング技術の構築、(b) 人流の特徴量を人間が理解できる形で意味づけするための汎用的な言語化技術の開発、技術 (a) (b) を含む情報学の技術により経済・社会活動をミクロな視点から広範囲に捉えられる (c) 人流に関する情報学と社会科学との融合基盤構築を下記のように進めた。

(a) 人流に関する汎用的なデータマイニング技術の構築は杉山グループが担当した。この研究項目では、人流ビッグデータから効率的に特徴量を抽出するデータマイニング手法の構築をおこなった。人流ビッグデータは高次元かつ大規模であり、そのままと既存の機械学習手法による解析が困難である。そこで、人流ビッグデータをグラフ構造データとして捉え、グラフマイニング技術を利用することで、人流ビッグデータからグラフの共通部分構造やグラフ間の類似度などのデータ解析に適した特徴量を抽出する、汎用的データマイニング手法を開発した。現在までに、グラフカーネルを用いた人流ビッグデータ処理系を確立し、グラフ構造を用いない単純な処理法と比較して、より精度の高いイベント検出が達成できることを確認した。

(b) 人流の特徴量を人間が理解できる形で意味づけするための汎用的な言語化技術の開発は宮尾グループが担当した。この研究項目では、人流の特徴量を人間が理解できる形で意味づけするための言語化技術を開発した。データの内容を説明するテキストを生成する技術は、画像のキャプション生成や株式市場の概況生成などさまざまな研究が行われているが、基本的には入出力ペアの大規模データを学習データとして機械学習を適用する。しかし、人流ビッグデータから抽出される特徴量は本質的に予測できないものであるため、既存のタスク設定では十分な精度が得られない。そこで、特徴量に加えて人流ビッグデータに紐付けすることができるテキストデータ（位置情報付きツイッター、時間・場所に紐付けられたニュース記事など）を利用する枠組みを検討した。現在までに、特定のイベントに紐づけられたツイート集合とそれを要約したテキストのデータセットを開発した。さらに、このデータセットを学習・評価データとして用いて、既存の自動要約手法の評価を行った。

(c) 人流に関する情報学と社会科学との融合基盤構築は水野グループが担当した。この研究項目では、経済・社会活動をミクロな視点から広範囲に捉えられる既存の人流ビッグデータ解析技術及び、本研究計画において構築する新技術の具体的な応用例を情報学と社会科学のコミュニティに提示することによって地球規模の課題解決に向けた融合基盤を構築するために、(c-1) 人流関連データの整備、(c-2) 情報学技術の社会問題への適用、(c-3) 相互理解のための研究成果の可視化、(c-4) 関連分野の研究者へのヒアリングをおこない、大型共同研究プロジェクト立ち上げに向けた連携を開始した。世界の主要都市における移動者の GPS 履歴データや、位置と個人属性の関係性データ、位置情報付きのツイートを整備し、近年、フランス・パリ近郊においてアラビア語ユーザーとフランス語ユーザーの空間分布に乖離が生じていることを示した。また、苗字・国籍ビッグデータから民族を推定する分類機を作成し、

地域の民族の空間分布を把握できるようにした。情報学、政治学、経済学の約 15 名の研究者へのヒアリングをおこない、大型共同研究に向けたいくつかの共同研究を立ち上げた。

(2) 顕著な成果

査読有り学会論文“全珠美, 水野貴之, 苗字・国籍ビッグデータによる民族の空間的特徴の把握, 計測自動制御学会第 18 回社会システム部会研究会プロシーディング, 2019”では, Recurrent Neural Network に, 全世界 3700 万の主要経済人(会社役員, 株主, 公的人物など)の苗字と国籍を学習させることにより, 国別に苗字生成の言語モデルを構築し, 苗字から民族(その苗字の起源なる国)を推定する分類器を作成した。構築された苗字の言語モデルに経済人依存はなく, 一般に適用できる。主要経済人の出生地や居住地の情報を用いて, 各地域の経済界における民族の空間分布を分類器により推定でき, 分布のエントロピーによって民族多様性を定量評価できるようになった。地域間での民族分布の比較により, 地域間の民族の類似度をネットワークで可視化することで, 空間的に広がる民族の特徴を把握できる。本成果を, 各年の個人電話帳や地域情報付き各種名簿データに応用することで, 年単位で地域の民族の人流分析が可能になることが期待できる。

<基礎研究としての成果>

これまで, スマートフォンなどの個人が持ち歩く通信機器や車載機器から収集された位置情報は, 都市設計や防災, 観光分野で主に研究や実務利用がされてきた。昨今, このような位置情報が国際政治と関係する社会問題の解決, 特に難民や不法移民の管理に利用されるようになってきた[ドイツ連邦移民難民局 2017]。これらは喫緊の課題であるために, 社会問題の解決のための科学的な位置情報分析の知見は不十分ではあるが, 実用化が既に始まっている。本研究プロジェクトは, この科学的知見の不足を解消するために, 情報学と社会科学の共同研究基盤を構築することであり, そのためのデータセットや国内外の研究者ネットワークの構築, 基盤技術の開発をおこなった。これらの取り組みは, 1 年 4 ヶ月という短い期間ではトップジャーナルの採択までは到達していないが, 来年度・再来年度に大きな成果を生み出す土台は十分に整った。本取り組みに関しては, 情報学と社会科学の各関連するコミュニティにおいて高い期待が寄せられている[Asian Network of Complexity Scientists Meeting 2019]。

§ 2 研究実施体制

(1) 研究チームの体制について

① 水野グループ

研究参加者

氏名	所属	役職	参加時期
水野 貴之	国立情報学研究所 情報社会相関研究系	准教授	H29. 12～
中渡瀬 秀一	同上	特任研究員	H29. 12～H30. 3
小高 充弘	同上	リサーチ・アシスタント	H30. 4～

研究項目

- ・ 人流に関する情報学と社会科学との融合基盤構築
- ・ 大型グラントの研究提案体制の構築

② 杉山グループ

研究参加者

氏名	所属	役職	参加時期
杉山 鷹人	国立情報学研究所 情報プリンシプル研究系	准教授	H29. 12～

研究項目

- ・ 人流に関する汎用的なデータマイニング技術の構築

③ 宮尾グループ

研究参加者

氏名	所属	役職	参加時期
宮尾 祐介	東京大学大学院情報理工学系研究科	教授	H29. 12～
近藤 崇宏	同上	技術補佐員	H29. 12～

研究項目

- ・ 人流の特徴量を言語化する汎用的な技術の構築

(2) 国内外の研究者や産業界等との連携によるネットワーク形成の状況について

本プロジェクトでは「大型グラントの研究提案体制の構築」するために、国内外の情報学と社会科学の研究者に本プロジェクトに関するヒアリング調査をおこなった。そして、本

プロジェクトを土台とした大型プロジェクトを立ち上げるための国内外の分野融合型共同研究を開始した。ここでは、公開可能な新たにスタートさせた共同研究について記載する。

1. 水野貴之 研究代表者。複雑なグローバル関係を紐解くデータ中心科学と国際政治学との分野融合の創出。国際共同研究加速基金(国際共同研究強化(A))，2018年度採択課題。この共同研究では、国際政治学者であるニューヨーク大学政治学部・准教授及び同大学データ科学センター・副所長のArthur Spirling氏とともにデータ科学と国際政治学との分野融合研究について、データ科学と国際政治学の両分野で国際的にリードする。
2. 水野貴之 研究分担者(研究代表者：栗崎周平)。ビッグデータで切り拓く国際関係学。基盤研究(A)。2018年度採択課題。この共同研究では、国際政治学者である早稲田大学の栗崎准教授とともに、国際政治分野におけるデータ科学の利活用を推進する。空間的セグリゲーションについて扱っており、本プロジェクトにおける「人流に関する汎用的なデータマイニング技術の構築」の成果を国際政治分野に応用している。
3. 水野貴之 研究分担者(研究代表者：渡辺努)。対話型中央銀行制度の設計。基盤研究(S)。2018年度採択課題。この共同研究では、マクロ経済学者である東京大学の渡辺教授とともに、マクロ経済分野、特に金融政策におけるデータ科学の利活用を推進する。市場の反応の特徴量を言語化することを扱っており、本プロジェクトにおける「人流の特徴量を言語化する汎用的な技術の構築」の成果をマクロ経済分野に応用している。