

平成 30 年 6 月 4 日現在

機関番号：14401

研究種目：挑戦的萌芽研究

研究期間：2016～2017

課題番号：16K12533

研究課題名(和文) 広報情報・オープンデータ・ソーシャル情報の融合による地域課題の横断的分析基盤

研究課題名(英文) A Cross Sectional Analysis Platform for Local Problems based on Integration of Public Information, Open Data and Social Information

研究代表者

古崎 晃司 (KOZAKI, Kouji)

大阪大学・産業科学研究所・准教授

研究者番号：00362624

交付決定額(研究期間全体)：(直接経費) 2,600,000円

研究成果の概要(和文)：本研究では、広報情報、オープンデータ、ソーシャル情報という異なる情報を組み合わせることにより、地域課題に分析する情報技術を開発した。具体的には、1)自治体の広報情報を自動収集して分析用の知識ベース(LOD)を構築するシステムの開発、2)自治体が公開している情報やオープンデータを基に、クラウドソーシングを用いて社会課題に関する知識ベース(LOD)を構築する技術の開発、3)人の移動履歴とオープンデータを組み合わせた分析事例の開発、を行った。

研究成果の概要(英文)：We developed information technologies which combine different kinds of data such as public information, open data and social information so that the users analyze local problems. The results include 1) a system automatically collects public information by local governments and construct a knowledge base (LOD) for data analysis, 2) a method to develop a knowledge base (LOD) for social problems based on public information and open data by local governments using crowd sourcing, and 3) a case study of data analysis based on combinations of GPS trajectory data and open data.

研究分野：オントロジ 工学

キーワード：オープンデータ オントロジー 地域課題 ソーシャル情報 広報情報

1. 研究開始当初の背景

誰でも自由に利用可能なオープンデータの公開・活用により、様々な社会問題の解決や新産業の創出を目指す取り組みは、2013年G8サミットでの「オープンデータ憲章」への合意など、世界的に注目されている。オープンデータ活用に関するコンテストや、市民参加型のイベントも多数開催されており、地域問題を情報技術で解決しようとするCivic Techと呼ばれる取り組みも各地で行われている。

このような地域課題の解決は、市民生活の直結する重要な課題であり、その地域の地理的特性、文化的・歴史的背景、産業構造、住民の特性など、様々な要因が関係しており極めて多様性が高い。このような多様性のため、地域課題に対する取り組みの実態を把握することは極めて難しく、「地域課題の解決に関する取り組みを継続的、かつ、地域横断的に分析・評価する仕組み」は十分に確立されていない。

2. 研究の目的

本研究では、地域課題を地域横断的に分析・評価する仕組みが確立されていないという問題を解決すべく、地域課題に関わる行政情報を発信している自治体の「広報情報」、地域の基本的な特性が現れる各種の「オープンデータ」、市民の声が現れる「ソーシャル情報」の3種の異なる情報を融合させることにより、地域課題を横断的に分析できる情報基盤を開発する。

これにより、全国の自治体が直面している地域課題を、それぞれの地域の特性を考慮した上で、横断的に分析することが可能となり、その実態の把握や解決策の検討に大きく貢献することを目指す。また情報技術的には、特定領域に注目した異データ融合型のデータ分析の基盤技術となることを期待する。

具体的には、以下の項目に沿って「全国の地域課題を横断的に分析する情報技術」を開発する。

(1)自治体の「広報情報」を自動収集し、関連する「オープンデータ」および「ソーシャルメディア」の情報と融合させ、それらの情報の横断的な検索ができる統合データベースを構築する。

(2)構築した統合データベースを用いて、地域課題を全国横断的に比較・分析する技術を開発する。

(3)(1)および(2)の技術を全国1,700市町村の9割以上を対象に適用し、「地域課題の横断的分析基盤」として公開する。

3. 研究の方法

(1)自治体の「広報情報」については、自治体がWebサイトで公開している更新情報を自動収集し、データベースに格納するシステムを開発する。図1に収集システムの全体像を示す。本システムは、Webサイトの更新情報

を記述する標準規格であるRSSにより新着情報を配信している自治体のWebサイトを対象とすることで、効率的な情報収集を行う。収集したRSSは、その内容を分析し想定対象者読者やカテゴリ、関連キーワードなどの追加情報を自動付与した上で、RDFデータベースに格納し、Webの標準技術に基づいて効率的な利用がしやすい形式のオープンデータであるLOD(Linked Open Data)として公開する。公開されたLODは、ユーザの関心に応じた情報は情報配信や、自治体を横断したデータ検索・分析に利用できる。

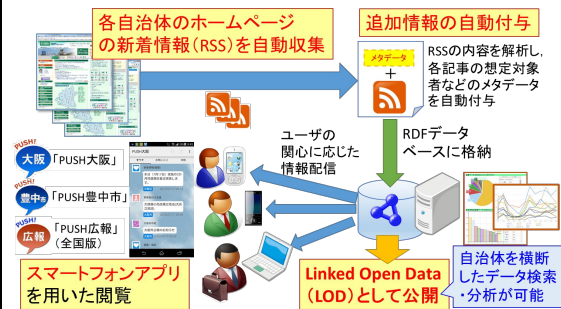


図1 広報情報の収集システムの全体像。

(2)オープンデータおよびソーシャル情報については、主に大阪市を対象として収集する。大阪市ではオープンデータを公開するポータルサイトを運営しており、APIを通じたオープンデータの取得が行える。また、ソーシャル情報については、TwitterのAPIを用いて、大阪市に関連するツイートを収集する。

(3)(1)と(2)の手法で収集した情報を統合することで、社会課題を分析する手法を開発する。分析手法としては、社会課題に関する知識を体系化したオントロジーや既存のLODを利用した知識処理的な手法と、ソーシャル情報の分析手法を応用したアプローチの適用を検討する。

4. 研究成果

(1)「広報情報」について、図1で示した全体像に沿って、自治体がホームページで公開している「新着情報」を自動的に収集し、データ分析用のメタデータ(更新日、情報の分類など)を付与した上で、データベースに格納するシステムを開発し、運用を開始した。本システムでは、収集した「広報情報」を、オープンデータから収集した社会課題に関するキーワードとマッピングした上で、Web技術として標準化されているクエリ言語SPARQLを用いた解析が可能なRDFデータベースに格納し、LODとして公開している。

データ収集の対象は、当初の研究目的では、全国自治体の9割を対象とすることを目指していたが、データ収集・分析の仕組みの開発に注力することを優先したため、現状の収集対象は約100自治体となった。

収集した広報情報は、(2)で述べるデータ分析に用いられると共に、市民が自分の興味

に応じて広報情報を受け取ることができるスマートフォンアプリケーション(「PUSH 大阪」「PUSH 公報」等)でも利用される。

また、2017年度末から、本システムによる広報情報の自動収集・LOD化の仕組みと、大阪市が市民活動向けの情報を発信するために運用しているポータルサイト(<http://kyodo-portal.city.osaka.jp/>)との相互連携を開始した。

(2)収集・公開した広報情報の LOD を解析する Web システムを開発した本システムでは、カテゴリ毎、自治体毎の配信記事数の統計といった基本的な比較(図2)に加え、月ごとの記事数の推移など、より詳細な分析に必要な基本機能を提供している。さらに詳細な分析が必要な際には、SPARQL による検索を用いることができる。

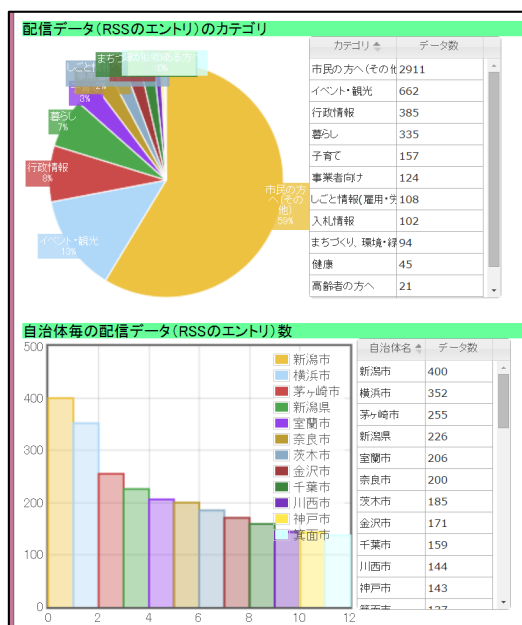


図2 配信記事数の比較機能。

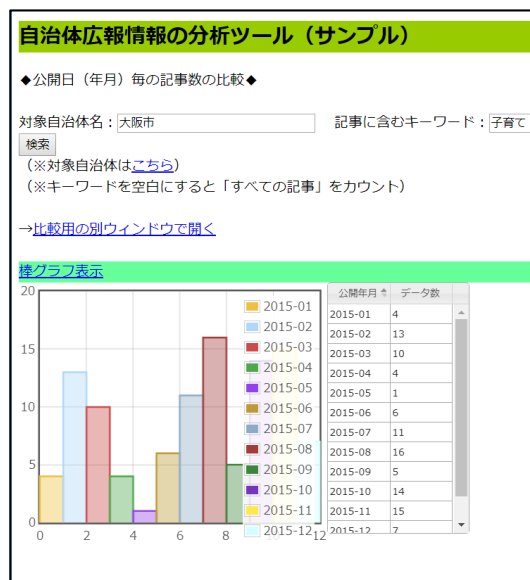


図3 月ごとの記事数の推移の分析。

例えば、自治体ごとに記事に多くあらわれるキーワードを比較すると、各自治体が力を入れている政策の特徴が伺える(表1)。

表1 記事に現れるキーワードの比較(対象データは2015年1月1日から2月1日にRSSで配信された記事)。

大阪市	横浜市	室蘭市	金沢市
キーワード 件数	キーワード 件数	キーワード 件数	キーワード 件数
施設 166	障害 50	臨時職員 89	気象 167
地域 159	保育所 50	建設 83	牛 118
福祉 154	アルバイト 41	定時 77	科学 55
住宅 149	福祉 38	水道 60	美術館 52
介護 91	企業 35	空き家 41	美術館 40
環境 86	介護 30	保育所 31	議会 37
消防 84	大学 30	住宅 29	大学 32
保育所 84	市営 28	議会 22	文化 32
教育 81	地下鉄 24	購入 22	乳幼児 22
法人 80	生活 20	市営 20	感染症 19

※「事業」など一般的なキーワードは除外している。

(3)「ソーシャルメディア」の情報の収集については、広報情報とオープンデータ収集のテストケースとして用いている大阪市に関連するツイートデータをTwitterから収集した。しかしながら、収集したツイートには大阪市の政策や社会課題に関する言及をしたものは、ごく少数しか含まれていなかったため、ソーシャル情報の収集方法のクラウドソーシングを利用した手法に変更した。

具体的には、自治体が公開している行政情報に関する文章、オープンデータ、Webから収集した地域課題に関する文書を組み合わせた情報を、クラウドソーシングを用いて地域課題の因果に関わる知識をLinked Open Data(LOD)化する手法を開発した。この手法を用いて構築したLODを評価するため、大阪市の行政職員および大阪市内で社会課題に関わる活動を行っている市民団体メンバーの協力を得て、ヒアリングによる評価を実施した。その結果、構築されたLODには、社会課題の理解を促進や議論の深化に有用な情報が含まれていると評価された。

(4)自治体が公開しているオープンデータと、ソーシャル情報を統合する別の活用事例として、GPSを用いて収集したセンサーデータをオープンデータから得られた情報と融合して分析するシステムを開発した。GPSデータはスマートフォンなどからも収集することが可能であるが、利用者のプライバシーを考慮して、個人情報が記録されないGPS専用の記録デバイスを利用し、事前に収集するデータ内容に同意を得た国際会議参加者の協力を得て位置データのみを収集した。収集した位置データは、場所に関する複数のオープンデータと統合し、協力者の名前を匿名とした上でLODとして公開した。公開したLODのデータ構造の概要を図4に示す。本データは、人が改心を持つ場所に関するPOI(Point Of Interest)情報と、実際に人が移動した場所の履歴を表している。これにより、誰が、いつ、どの場所を、どのような順番で移動したかを分析することができる。

POI情報は既存のオープンデータから得た

施設情報、観光地、駅など、の情報をを用いている。よって、緯度経度による数値データを、施設名に変換した上で移動履歴を分析できる。分析結果の可視化例を図5に示す。

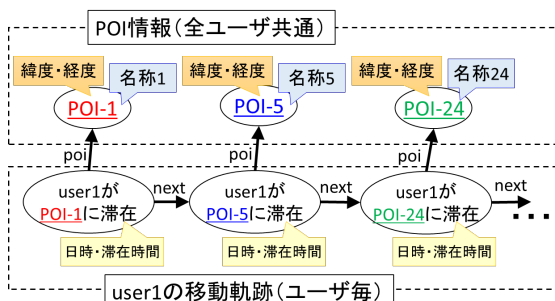


図4 移動履歴のLODの概要。



図5 移動履歴のLODの分析例。

また、このような分析にはPOI情報が充実していることが重要であるが、本研究では複数のオープンデータを利用していることから、それぞれのオープンデータの特徴比較としての意義も持つ。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計5件)

熊澤 輝一、古崎 晃司、環境・サステイナビリティ領域におけるドメイン知識間の因果論理構築支援ツールの開発、人工知能学会論文誌、査読有、33巻、2018、E-SGAI04_1-13
DOI: 10.18919/jkg.67.12_633

古崎 晃司、横山 輝明、深見 嘉明、GPS移動履歴の収集とオープンデータを用いた移動軌跡のLOD化 国際会議 ISWC2016における実証実験を例として、デジタルプラクティス、査読有、9巻、2018、138-163

古崎 晃司、コミュニティ活動を通じたLOD活用の“つながり” - LODハッカソン関西を例として -、情報の科学と技術、査読無、67巻、2017、633-638

DOI: 10.18919/jkg.67.12_633

増田 壮志、古崎 晃司、概念階層の比較に基づくオントロジー品質向上支援手法、人工知能学会論文誌、査読有、32巻、2017、E-G71_1-10

DOI: 10.1527/tjsai.E-G71

[学会発表](計20件)

Shusaku Egami、Linked Urban Open Data Including Social Problems' Causality and Their Costs、7th Joint International Semantic Technology Conference (JIST2017)、2017年、ゴールドコースト(オーストラリア)

Kouji Kozaki、GPS Trajectory Linked Open Data based on Open POI Information- Through an Experiment in ISWC2016-、16th International Semantic Web Conference (ISWC 2017)、2017年、ウィーン(オーストリア)

古崎 晃司、Linked Data とオントロジーによるセマンティック技術の実際、2017年度人工知能学会全国大会(第31回)(招待講演)、2017年、名古屋(愛知県)

古崎 晃司、国際会議参加者の移動履歴の収集とLOD化の試み、第41回セマンティックウェブとオントロジー研究会、2017年、福岡市(福岡県)

古崎 晃司、地域課題分析に向けた自治体広報情報の収集とLOD化の試み、第38回セマンティックウェブとオントロジー研究会、2016年、宮古島(沖縄県)

古崎 晃司、自治体広報情報のRSSに基づく地域課題分析の試み、2016年度人工知能学会全国大会(第30回)、2016年、北九州市(福岡県)

[その他]

広報情報のLOD(PUSH大阪&PUSH公報)公開サイト、<http://push.jp.net/>

移動履歴のLOD(GPS2LOD)公開サイト、<https://github.com/koujikozaiki/GPS2LOD>

6. 研究組織

(1)研究代表者

古崎 晃司 (KOZAKI, Kouji)
大阪大学・産業科学研究所・准教授
研究者番号: 00362624

(2)研究分担者

川村 隆浩 (KAWAMURA, Takahiro)
電気通信大学・大学院情報理工学研究所
・客員准教授
研究者番号: 10426653