

特集投稿論文

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—

古崎 晃司¹ 横山 輝明² 深見 嘉明³

¹大阪大学 ²神戸情報大学院大学 ³立教大学

国内の経済活性化のためインバウンド観光の振興が期待されている。そこでICTを活用した支援として、観光客の位置情報を収集し、訪問場所の分析する取り組みが始まっている。ところがGPSなどのセンサから生成される生データは大量となりデータ分析は負担となる。そこで本稿では、オープンデータから取得したPoint of Interest (POI) 情報を利用して、GPS移動履歴をRDFによるグラフデータに変換しLinked Open Data (LOD) 化することで、柔軟なデータ分析を容易にする手法について報告する。本手法の実証実験を神戸にて開催された国際会議ISWC2016において実施し、LOD化によりSPARQLクエリを用いた容易な分析が可能であることと、既存オープンデータのPOI情報として有用性を確認した。

1. はじめに

1.1 国際会議参加者の移動履歴の分析

2008年に観光庁が設立され、国外からの旅行者、すなわちインバウンド需要創出による経済活性化が国策として進められている。観光庁において重点的な施策とされているうちの1つがMICE需要の拡大である。MICEとは、M (=Meeting) ・ I (=Incentive tour) ・

C (=Convention) ・ E (=Exhibition/Event) の頭文字からなる造語であり，国際会議や企業の従業員に対する報奨旅行，国際展示会といった大規模会場を活用した商業イベントを総称するものである。2013年閣議決定の日本再興戦略[1]においても「2030年にはアジア No.1の国際会議開催国としての不動の地位を築く」という目標が掲げられている。観光立国実現に向けたアクション・プログラム[2]においても，観光立国実現に向けた主要な柱の1つとしてMICEが位置付けられている。

2012年3月に策定された観光立国推進基本計画[3]においては，「今後5年以内に我が国における国際会議の開催件数を5割以上伸ばし，アジアにおける最大の開催国を目指す」との目標が掲げられた。2016年3月に策定された明日の日本を支える観光ビジョン[4]においても，MICE誘致は有力な施策として取り上げられ，観光庁による図1のようなロゴも策定されている。

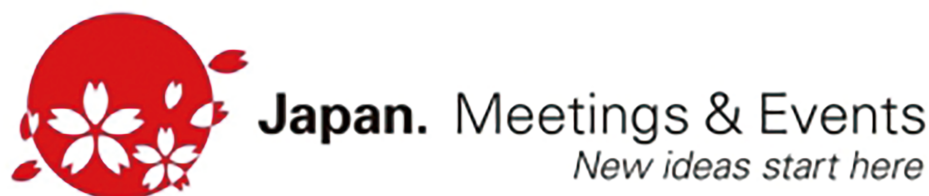


図1 観光庁が定めたMICE誘致ロゴ
(<http://www.mlit.go.jp/kankocho/shisaku/kokusai/mice.html>より引用)

観光庁の調査[5]によると，MICEの需要は世界的には拡大基調であり，2006年から2015年の間に1.4倍の開催件数となっている。大会ごとの参加人数は減少トレンドであるものの，件数の拡大によってカバーされており，MICE市場の重要性は高まっている。日本の世界におけるポジションは，同調査によると2015年の国際会議開催件数では世界7位・アジア1位となっている。ただし，米国等の上位国との差は拡大傾向にある。また，2014年まで継続してアジア1位の座にあった中国との差も小さい。

1.2 従来のMICE振興施策と都市競争力

このように国外MICE需要の取り込みは重要な課題とされており、さまざまな施策が講じられている。その1つがユニークベニユーの活用／創出である。ユニークベニユーとは、「歴史的建造物，文化施設や公的空間等で，会議・レセプションを開催することで特別感や地域特性を演出できる会場」[6]と定義されている。

ユニークベニユーの例としては沖縄の美ら海水族館などがある。しかし，MICE参加者は日本滞在中に会場施設内に留まり続けるとは限らない。日本平ホテルのような市街地から離れた会場であればともかく，都市に所在するベニユーで開催されるイベントの場合，参加者は空き時間，もしくは開催日前後の日程で周辺市街地にある観光スポットを回遊する。MICE開催地としての競争力は会場単体ではなく，会場を取り巻く都市全体のもつ魅力の総体であるといえよう。MICE参加者がどのように街を回遊するか，またどのようなコンテンツが回遊を生み出し，滞在体験を向上させるのか，という視点がMICE誘致における都市の競争力醸成にとり重要なのである。

1.3 オープンデータを用いた移動履歴の分析とLOD化

インバウンドMICE振興のためには，ユニークベニユーなど魅力的な会場を発掘する以外にも，開催都市を面として捉え，都市内を回遊するという観点からの魅力発掘や開発が必要となってくる。そのため，来訪者がどのような観光スポットを回遊するかについて実態を分析する必要がある。

来訪者の行動分析のために観光ビッグデータ解析が実施されている。観光庁による，GPSを利用した観光行動の調査分析[7]では，観光ビッグデータを「観光客の観光行動を反映する大規模・多種・複数情報源由来のデータ群」と定義している。具体的にはスマートフォン等を利用して通信回線を利用した際の基地局情報，位置情報や具体的な場所の名称を含んだSNS投稿，GPS機能による位置情報などが用いられている。

平成27年度ICTを活用した訪日外国人観光動態調査[8]では，

- ・携帯電話の基地局情報が，広域での観光による来訪者の集積状況分析に
- ・SNS投稿の情報が，観光スポットの訪問目的や評価分析に

・GPSデータが、移動経路や集積ポイントなどの地域におけるマイクロの動態やスポットの吸引力分析に

用いられている。

この中でGPSデータが都市内の動態分析に活用されているが、GPSデータのみを用いて分析を行うという手法には限界がある。取得されるデータがGPSによる緯度経度情報のみであるため、具体的なスポット、つまりはPoint of Interest (POI) 情報との紐付けは事後かつ人為的に行うこととなるからである。

あらかじめ緯度経度情報に紐付けられたPOI情報を用意しておき、それとGPSデータをあわせて分析することができれば、データ解析時の手順が削減されるばかりではなく、より精緻な動態分析が可能となる。

POIデータは、できるだけ多様な観点で生成された多様なものを統合して用いることが、精緻な分析を実施するためには望ましい。つまり、自由に利用できる多様なPOIデータが供給されることがMICE振興の観点からの都市間競争には重要となる。これは、多様な主体が標準的な仕様に則ったオープンデータとしてPOI情報を公開することによって実現される。すなわち、多様なデータ資源がオープンデータとして提供されることを前提とした効果的な分析手法の開発が、MICE振興施策の観点からの都市間競争力醸成のために望まれているといえる。

本研究では、このような背景を踏まえ、オープンデータを用いて収集したPOI情報に基づき、GPSデータを分析する手法の開発と、実データを用いた分析の実践を目的としている。GPSデータの分析には、Webでのオープンデータ公開のための技術仕様としてW3Cが策定しているLinked Open Data (LOD) を用いる。具体的にはGPSデータ（移動履歴情報）とオープンデータから収集したPOI情報を統合し、LODとして公開することで、さまざまな観点からの分析を可能とする。

1.4 提案手法の概要

図2に本研究で提案するGPSデータをLOD化する手法の概要を示す。本手法では、収集したGPSデータをLODの公開に用いられるデータモデルの標準規格であるRDF（Resourced Description Framework）^{☆1}に準拠したデータ形式に変換する。その上でRDFデータをLODとして公開することで、LOD標準のWeb APIであるSPARQLエンドポイントを用いたデータ分析を可能とする。LOD化にはオープンデータとして公開されているPOI情報を利用し、GPSログから得られた移動軌跡を分析しやすいように設計したRDFデータモデルに基づいて変換する。本研究ではこれらの手法の中でも特に、

- ・ GPSデータの収集方法
- ・ POI情報として用いるオープンデータの収集
- ・ GPS 移動軌跡のためのRDFデータモデルの設計

に注力する。

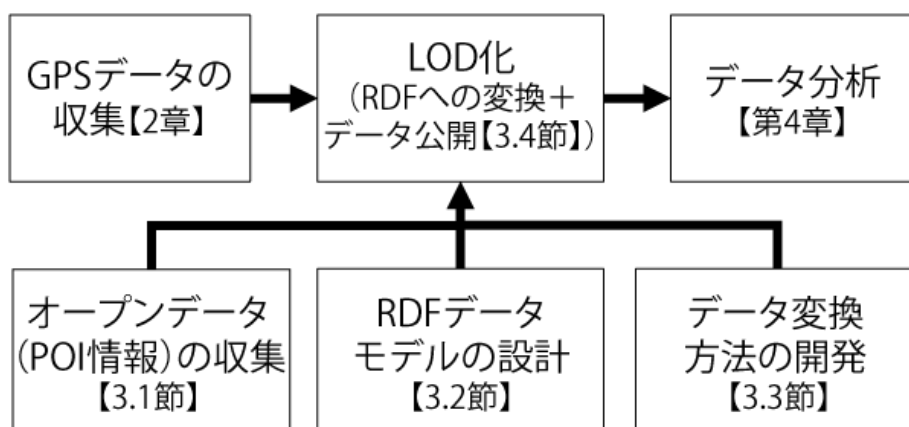


図2 オープンデータを用いたGPSログのLOD化の手法

以下、本稿は次のように構成される。第2章では、GPSデータの収集方法と国際会議ISWC2016の参加者を対象に行った移動履歴の収集実験について述べる。第3章では、GPSデータをLOD化する手法とともに実際に収集したデータをLOD化した結果を述べる。第4章ではLOD化したGPS移動軌跡の分析例を示し、続く第5章で本稿を総括するとともに今後の課題を述べる。

2. 移動履歴の収集実験

インバウンド来訪者の都市内回遊などの行動を知るために、来訪者の移動情報の収集が必要となる。スマートフォンのようなGPS機能を搭載した高機能端末も一般化しており、人々の移動情報の自動収集が可能となっている。ここでは、あるエリア内での人々の関心や意図などを明らかにするための移動履歴の収集実験の方法について、既存手法との比較を含めて検討したのち、本研究における移動履歴の収集実験の結果について報告する。

2.1 移動履歴の収集方法の検討

情報通信技術の発達に伴い、ユーザの移動情報を自動収集するためのさまざまな方法が登場している。ユーザ位置を推定するための代表的な方法として、(1) 自律測位、(2) 計測センサ利用の2種の方法について説明する。

Global Positioning System (GPS) による自律測位は、位置情報取り扱いにおいて主流の技術である。GPSでは30個超の人工衛星の位置と人工衛星からの電波信号の遅延に基づく距離を利用した三角測量にて、端末位置を推定する。そのため、空が見える場所であれば世界中で利用可能という利点を持つ。しかし電波信号に依存するため、ビルなどの遮蔽により測位誤差が発生すること、電波受信や測位計算のために端末には演算機能が求められ、端末の電源確保が必須となる、端末の実装コストがかかるなどの欠点も持つ[9]。GPSはスマートフォンの標準的な機能として実装されたため、広範に利用されることとなった。

対して、環境側に設置したセンサにより近傍ユーザを識別する形態でのユーザ位置の推定手法もある。対象の検知にはさまざまな方法が利用されており、典型的な例はRFIDなどの無線タグに対する検知ゲートである。この方法では、ゲート通過の検知によりタグを付与された対象が当該時刻にゲートにいたことを推定する。ほかにも、カメラ利用[10]、スマートフォンからのWiFi電波を利用[11]するなどさまざまなセンシング方法が提案されている。これらの手法は、GPSなどの電波基準局を用いる方法と比べると、位置を検知できる場所がセンシング場所に限定されるというカバー範囲が局所的

になることが欠点である。しかし、検知はほぼ確実にその場所にいたことを意味する情報の正確さ、検知対象側に複雑な機器を必要としないことでのコスト安や堅牢性という利点も持つ。

自律測位と計測センサの両方式は、端末コスト、カバー範囲においてトレードオフの関係を持つ。自律測位では面的なユーザ追跡となり、計測センサでは点でのユーザ追跡となる。対象の時間と位置を確定させるという観点では、空間分解能が異なる同種の情報として取り扱うことも可能である。

これらの技術要素が整ったことで、位置情報の収集や応用は一般的となり、大量のユーザの移動情報について収集分析が期待されており、その期待とプライバシー懸念が議論されている[12]。位置情報の持つ社会的有用性を活用するために、匿名化や統計処理などのプライバシー対策を施したガイドラインが示されている。

これらの技術的背景と社会的背景より、本実験での利用技術や実験実施方法を検討した。学会参加者を対象とした実験として被験者が限定的であること、被験者側での操作等のオペレーションの手間を軽減させることを考慮し、小型のGPS端末を配布・所持してもらい計測することとした。またプライバシー対策として、被験者の個人情報収集せず、得られた移動履歴と特定個人との対応の情報は残らないよう配慮した。

2.2 移動履歴収集の実験計画

インバウンド来訪者の都市内回遊を把握するための分析の実験として、国際学会関係者を来訪者とみなした実証実験を実施する。分析手法となるLODとも関連の深いInternational Semantic Web Conference 2016 (ISWC2016)を実験対象とし、神戸で実施される本学会参加者を被験者として、神戸における回遊を分析した。約400名の学会参加者へ事前の実験協力を依頼したところ約20名からの協力の申し出があり、用意できた端末の台数の制限から先着順に端末を配布し、最終的には11名分のデータが収集できた。

協力者にはあらかじめ学会内の端末配布場所を伝えておき、協力者はISWC2016参加の初日にGPS端末を受け取る。受付側では協力者に関する情報は保持せずに端末配布を実施した。データ収集の簡

素化のため端末の電源は常時オンとして、協力者に対して端末を保持している間は常時、位置情報が記録される旨を説明した上で、会議の会期中、位置情報の記録に差し障りがない時間帯にデバイスを携帯して移動することを依頼した。実験手順の簡素化と操作トラブル防止のため端末の電源操作はできない状態で配布した。すなわち、位置情報の記録を望まない場合は、端末を携帯しないことで記録の制御を協力者側に委ねる。配布端末と協力者は紐づけておらず、個人の追跡はできないなどのプライバシー配慮に基づいて実験を実施した。

GPS端末には、市販のi-gotU GT-600を使用し、データの記録は約1分ごとの位置情報（緯度、経度、標高、移動速度、移動距離）および日時・時刻とした。同端末では1分ごとの位置情報の記録設定であれば、会期中の5日間に充電不要で動作するため、被験者に追加作業を求めない実験実施が可能であった。また端末はネットワーク接続機能を有しておらず、データは端末より直接読み込む必要がある。よって会議の最終日に協力者からGPS端末を回収し、データを収集した。

2.3 移動履歴収集の結果

表1に会期中の2017年10月17日-21日に11名から収集したデータの概要を示す。GPS端末は電波状況の悪化などで測位が不能な場合にはデータを記録しない。そのため、得られたデータ数は測位できたタイミングのものに限定された。最大移動距離は収集したデータの軌跡の移動範囲である。学会会場を起点として神戸市近辺での移動が観測されていることがわかる。このような結果から、想定通り、会期中を通じて十分なデータ収集が可能であったといえる。

表1 収集したGPSログデータの概要

被験者	記録開始(※)	記録終了(※)	データ数	記録日数 (日)	記録時間 (時間)	最大移動 距離 (m)
user1	10/17 11:18	10/21 16:53	633	4.23	102	6,914
user2	10/17 11:00	10/21 16:46	320	4.24	102	6,050
user3	10/18 11:01	10/21 14:00	315	3.12	75	4,284
user4	10/18 11:02	10/21 13:14	484	3.09	74	23,668
user5	10/18 10:59	10/21 16:33	386	3.23	78	3,646
user6	10/18 10:58	10/21 16:48	354	3.24	78	3,650
user7	10/17 11:22	10/21 15:04	511	4.15	100	4,656
user8	10/17 10:59	10/21 13:29	309	4.10	98	43,555
user9	10/17 11:32	10/21 16:50	433	4.22	101	4,066
user10	10/17 11:25	10/21 16:49	170	4.22	101	6,531
user11	10/17 11:30	10/21 16:47	193	4.22	101	3,663
平均	-	-	373.5	3.83	91.8	10,062
中央値	-	-	354.0	4.15	99.7	4,656

※記録開始・終了の日時はいずれも2016年10月

3. GPS移動軌跡のLOD化

第2章で述べた手法により収集したGPSの移動履歴データは、オープンデータから取得したPoint of Interest (POI) 情報、および、移動軌跡を表現するRDFデータモデルに基づいてRDF形式に変換され、LODとして公開される。本章では、このGPS移動軌跡のLOD化の手法について述べる。

3.1 POI情報取得に用いたオープンデータ

本研究におけるGPS移動履歴のLOD化は、国際会議の参加者を対象としており、主に観光の観点から分析を想定している。よって、分析に用いるPOI情報は、観光に関連する位置情報を含むオープンデータから取得する。

第2章で述べた方法で収集したGPSログデータは、user8を除き移動履歴の範囲が神戸市内であった。そこで今回のLOD化においては、POI情報の収集範囲についても神戸市内に限定し、以下の6種類のオープンデータを利用した。

・神戸市がオープンデータポータル^{☆2}で公開している観光関連のオープンデータのうち、位置情報を含むもの。具体的には(1)観光施設情報、(2)夜景スポット、(3)ロケ地、(4)野外彫刻、の4種を用いた。

・ (5) DBpedia Japanese^{☆3}および (6) Wikidata^{☆4}から抽出した、位置情報を含むデータ。データ取得の際には、それぞれのSPARQLエンドポイント（クエリによるデータ取得用API）を用いて、2.3節・表1のGPSログの移動履歴が含まれる範囲の緯度経度を含むデータを収集した。さらに、取得したデータから、位置情報が神戸市内となるもののみを抽出^{☆5}して利用した。

(1) - (4) は、今回のPOI情報の収集範囲とした自治体である神戸市が提供しているオープンデータであることから、その地域に密着した観光情報に関するPOIが取得できることが期待される。一方、

(5) および (6) は、ウィキペディアのようにWeb上で誰でも編集できる環境下で、観光情報に限定しない汎用的な知識として構築されたオープンデータである。そのため、神戸市のオープンデータと比較して幅広い種類のPOI情報が含まれることが想定される。

さらに、これら6種類のオープンデータを統合したデータセットとして、

- ・ 神戸市のオープンデータである (1) - (4) を統合したデータ
- ・ (1) - (6) のすべてのオープンデータを統合したデータ

を作成した。その際、異なるデータセットには同一のPOI情報が重複して含まれる場合があることを考慮して、“距離が50m以下にある異なるデータセットから得られたPOIは、1つのPOI情報としてまとめる”という処理を施した。

これらの結果、得られた神戸市内のPOI情報の件数を表2に示す。

(1) - (6) を統合したデータセットにおけるPOI情報の数は730、データセット内に含まれるPOI間の距離（間隔）は、平均234.3m、中央値で139.1mであった。

表2 オープンデータから収集したPOI情報

データセット名	POI数	データセット内のPOI間の距離 (m)			
		最小	最大	平均	中央値
(1) 観光施設情報	103	2.4	4,743.9	546.8	210.3
(2) 夜景スポット	22	75.7	14,104.0	1,469.0	531.2
(3) ロケ地	88	37.9	4,690.4	504.0	233.7
(4) 野外彫刻	35	13.0	157.1	46.4	35.5
(1) - (4) の統合	223	2.4	4,743.9	332.5	125.6
(5) DBpedia Japanese	178	11.1	2,305.9	419.4	281.1
(6) Wikidata	544	2.2	3,222.1	256.3	175.3
(1) - (6) の統合	730	2.4	4,743.9	234.3	139.1

3.2 移動軌跡のRDFデータモデル

図3にGPS移動軌跡のRDFデータモデルの概要を示す。本データモデルは、POI情報および各ユーザの移動軌跡のそれぞれを表す2種類のデータモデルから構成される。

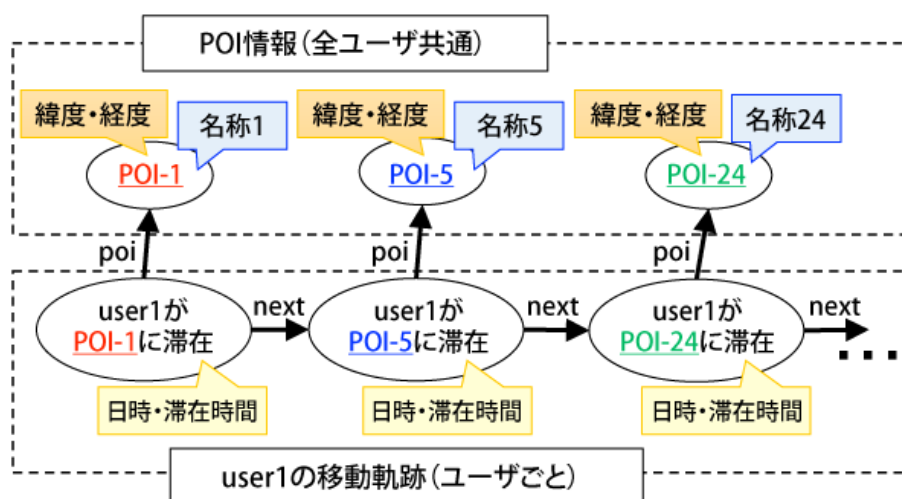


図3 GPS移動軌跡のRDFデータモデルの概要

RDFにおいてデータは、リソースと呼ばれる。すべてのリソースはURL (Uniform Resource Locator) の拡張仕様であるIRI (Internationalized Resource Identifier) を識別用のIDとして用いる^{☆6}。各リソース (データ) の内容は、「主語 (subject) -述語 (predicate) -目的語 (object)」の3つ組 (トリプルと呼ばれる)

の組合せで記述される。主語にはIRIで表されるリソースが、目的語にはリソースまたは文字列（データ型を指定することも可能）が入り、述語にはプロパティと呼ばれる関係が用いられる。よってRDFデータモデルの設計は、各データがクラスごとにどのようなプロパティを持つ（ことができる）かを定義することに相当する。

POI情報のデータモデルでは、POIクラスのインスタンス（リソース）において、POIの名称、位置情報（緯度経度）等を表3(a)に示すプロパティを用いて表現する。これらのPOI情報は、各ユーザの移動軌跡から共通に参照される。

表3 GPS移動軌跡のRDFデータモデルに用いるプロパティ

Prefix (接頭語) の定義

gtl : <http://lodosaka.jp/iswc2016gtl-exp/data/> .
 gtl-prop : <http://lodosaka.jp/iswc2016gtl-exp/prop#> .
 gtl-class : <http://lodosaka.jp/iswc2016gtl-exp/class#> .
 rdf : <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
 rdfs : <http://www.w3.org/2000/01/rdf-schema#> .
 xsd : <http://www.w3.org/2001/XMLSchema#> .
 geo : <http://www.w3.org/2003/01/geo/wgs84_pos#> .

(a) POIクラスのプロパティ

プロパティ名	値域	説明
rdfs:label	文字列	POI の名称
geo:lat	xsd:float	POI の緯度
geo:lon	xsd:float	POI の経度
gtl-prop:source	文字列	POI 情報の取得元のオープンデータ 複数の値を持つことが許される
rdfs:seeAlso	IRI	POI の関連情報への外部リンク (DBpedia Japanese, Wikidata へリンク) 複数の値を持つことが許される

(b) GPS移動軌跡のRDFデータモデルに用いるプロパティ

プロパティ名	値域	説明
gtl-prop:user	文字列	滞在したユーザの匿名化されたID
rdfs:label	文字列	滞在したPOIの名称
gtl-prop:poi	IRI	滞在したPO (POIリソースを参照)
gtl-prop:date	yyyy-mm-dd	POIに入った年月日
gtl:time	xsd:int	POIに入った時間 (0-23)
gtl-prop:start	xsd:dateTime	POIに入った日時・時刻
gtl-prop:end	xsd:dateTime	POIを出た日時・時刻
gtl-prop:next	IRI	次の滞在情報

一方、ユーザの移動軌跡のデータモデルでは、各リソースが「ユーザがあるPOI (スポット) に滞在した」という滞在情報 (StayPOIクラスのインスタンス) を表す。滞在情報には表3(b)に示すプロパティを用いて、ユーザ名、滞在したPOI、滞向日時・時間などを表現し、滞在したPOIについてPOI情報のリソースを参照して表す。なお、gtl:timeは「〇時以降」といった条件をクエリで指定する際に、gtl-prop:startでは「年月日+時刻」の指定が必要となり煩雑となるため、時 (何時か) のみを独立させたプロパティとして導入した (使い方は第4章・図5のクエリ例 (d) を参照)。

さらに、滞在情報に次の滞在情報へのリンク (nextプロパティ) を持たせることで、ユーザごとの移動軌跡が「滞在情報を頂点、次の滞在先へのリンク (移動を表す) を辺とする有向グラフ」として表現される。

図4の(a)および(b)に、上述のデータモデルに基づいて変換されたPOI情報 (神戸国際会議場) および滞在情報 (user1の移動軌跡における「神戸国際会議場への滞在」) のRDFデータの実例を示す。このPOI情報は、3つのオープンデータ (1) 観光施設情報、(3) ロケ地、(6) Wikidataから取得されたことがgtl-prop:sourceにより表されるとともに、Wikidata上における神戸国際会議場のIRIへのリンクがrdfs:seeAlsoを用いて示されている。このリンクを通して、こ

のPOIの関連情報を取得することができる。またこのPOI情報が、滞在情報のRDFデータから滞在したPOIとして参照されていることが gtl-prop:poiにより表されている。

gtl:poi-43

```

rdf:type gtl-class:POI;
rdfs:label "神戸国際会議場";
geo:lat "34.666234"^^xsd:float;
geo:lon "135.21301"^^xsd:float;
rdfs:seeAlso <http://www.wikidata.org/entity/Q11589835>;
gtl-prop:source "ロケ地";
gtl-prop:source "Wikidata";
gtl-prop:source "観光施設情報".

```

(a) POI情報のRDFデータ例：神戸国際会議場

gtl:user1-7

```

rdf:type gtl-class:StayPOI;
gtl-prop:user "user1";
rdfs:label "神戸国際会議場";
gtl-prop:poi gtl:poi-43;
gtl-prop:date "2016-10-17";
gtl-prop:time "12"^^xsd:int;
gtl-prop:start "2016-10-17T12:14:18Z"^^xsd:dateTime;
gtl-prop:end "2016-10-17T12:22:28Z"^^xsd:dateTime;
gtl-prop:next gtl:user1-8.

```

(b) 移動軌跡のRDFデータ例：神戸国際会議場への滞在

図4 GPS移動軌跡のRDFデータ例。Prefix（接頭語）は表3と同一のものを使用している

なおRDFデータにおいて、地域が単なる文字列ではなくxsd:floatやxsd:dateTimeといった型付きのデータの時、値にも型を明記する必要がある。また、RDFデータにプロパティと値の組を記載する順番には指定はなく、任意の順に入れ替えてもよい。

3.3 移動軌跡のRDFへの変換

GPSの移動履歴データ（以下、GPSデータと呼ぶ）のRDFへの変換は、CSV形式で保存したGPSデータを時系列に沿って1レコードずつ読み込み、下記の手順でPOIへの滞在を判定した後、3.2節で述べたデータモデルに基づきRDF形式で出力することで行った。

1.GPSデータの位置情報（緯度・経度）と、3.1節で述べたPOI情報に含まれる各POIの位置情報（緯度・経度）を比較し、

- 両者の距離が閾値d以下であり
 - その距離が他のPOIと比較して最小となる
- ときに、そのPOIに滞在したと判定する。

2.時系列が連続するGPSデータのレコードが同一のPOIに滞在すると判定された際、該当レコードのうち、

- 最も早い日時・時刻を、そのPOIに入った（滞在を開始した）日時・時刻
 - 最も遅い日時・時刻を、「そのPOIから出た（滞りを終了した）日時・時刻
- とする。

3.時系列順で、滞在するPOIが変化したとき、そのPOIを次の滞り先とみなす。

データの変換処理には、Javaを用いて独自に開発した変換プログラムを利用した。

3.4 LODとしての公開

収集した11名分のGPSデータを3.3節で述べた手法でRDF形式に変換した。変換時に用いるPOI情報には3.1節で述べた(1) - (6)のオープンデータを統合したものを利用し、POIの滞在判定に用いる距離の閾値 $d = 100\text{m}$ とした。変換結果として得られた滞在情報の数は1,462、移動軌跡のRDFデータの総トリプル数（POI情報のトリプルを除く）は13,158であった（2017年5月8日時点^{☆7}）。

これらのRDFデータは、GPSの元データ（GPX形式およびCSV形式）、POI情報、LOD化に用いたプログラム群のソースコードとともに、GitHub上で公開した^{☆8}。合わせて移動軌跡とPOI情報のRDFデータを格納したSPARQLエンドポイントを公開した^{☆9}。

4. LOD化したGPS移動軌跡の分析

LOD化したGPS移動軌跡の分析

第3章で述べた手法により，GPS移動履歴をオープンデータから取得したPOI情報を用いてLOD化したGPSの移動軌跡データの分析方法と事例について議論する．4.1節ではLOD用の標準クエリ言語であるSPARQLを用いた分析について，4.2節ではLOD化に用いるオープンデータのPOI情報としての有用性について考察する．

4.1 SPARQLクエリによる分析

本研究でLOD化したGPSデータは，ユーザ1人ごとの移動軌跡がデータ化されているため，個人単位と全体傾向の双方から分析できる．

分析には，ユーザ，POI（緯度経度，経路），時間（日時，滞在時間）など，表3で示したプロパティを任意に組み合わせたSPARQLクエリを用いる．たとえば，個人単位での分析例として，user1の移動軌跡の情報を時間順にすべて取得する際には，図5(a)のクエリを用いればよい．

```
PREFIX gtl-prop: <http://lodosaka.jp/iswc2016gtl-exp/prop#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
```

```
SELECT DISTINCT * where {
  ?s gtl-prop:user "user1";
    rdfs:label ?label;
    gtl-prop:poi ?poi;
    gtl-prop:date ?d;
    gtl-prop:start ?st;
    gtl-prop:end ?ed;
    gtl-prop:next ?g.
  ?poi geo:lat ?lat;
    geo:lon ?long.
}ORDER BY ?st.
```

(a) user1の移動軌跡の情報を時間順にすべて取得するクエリ


```
PREFIX gtl-prop: <http://lodosaka.jp/iswc2016gtl-exp/prop#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
```

```
SELECT DISTINCT ?label ?lat ?long
      (count(distinct ?u)AS ?c) where {
  ?s    gtl-prop:user ?u;
        rdfs:label    ?label;
        gtl-prop:poi  ?poi.
  ?poi geo:lat  ?lat;
        geo:lon  ?long.
  {GROUP BY ?label ?lat ?long
PREFIX gtl-prop: <http://lodosaka.jp/iswc2016gtl-exp/prop#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
```

```
SELECT DISTINCT ?u where {
  ?s1 gtl-prop:user ?u;
        rdfs:label " 生田神社 ".
  ?s2 gtl-prop:user ?u;
        rdfs:label "UCC コーヒー博物館 ". }
```

(c) 複数の地点の共に訪問したユーザを取得するクエリ

```
PREFIX gtl-prop: <http://lodosaka.jp/iswc2016gtl-exp/prop#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
```

```
SELECT DISTINCT ?poi ?label where {
  ?s    gtl-prop:user "user1";
        gtl-prop:poi  ?poi;
        gtl-prop:time ?t.
FILTER(?t >= 18)
  ?poi rdfs:label ?label. }
```

(d) user1が会議期間中の18時以降に訪問したPOIを取得するクエリ

図5 GPS移動軌跡のLODを分析するのに用いる基本的なSPARQLクエリの例

一方、全体傾向の分析例として、期間中に最低1回訪問したユーザの人数でPOIをランキングするクエリを図5(b)に示す。この結果を地図上で可視化すると図6のようになる^{☆10}。この結果を見ると、ISWC2016の会場である「神戸国際会議場」には全員が一度は立ち寄っていることに加えて、多くの人が訪問した個所は、会議場以外では三ノ宮周辺に集中していることなどが分かる。



図6 各POIも1度は立ち寄った人数の可視化例

さらに，“生田神社とUCCコーヒー博物館の両方を訪れた人数”のように，複数の地点を指定しての経路を元にした分析も，図5(c)のような簡単なクエリの組合せで行える．このように，GPSデータをLOD化した利点の1つは，SPARQLクエリにより柔軟なデータ分析が行える点にある．

なお，ユーザが訪問した時間帯によって結果を絞り込みたい場合には，図5(d)のように，`gtl-prop:time`プロパティの値を取得し，絞り込み条件を記述するFILTER構文を用いて条件としたい時間帯を指定すればよい．より詳細な日時や時刻を指定したい場合は，訪問した年月日を値として持つ`gtl-prop:date`，年月日および時刻を値として持つ`gtl-prop:start`と`gtl-prop:end`プロパティを用いればよい．

これらの基本的な分析に加え，LOD化の最も重要な利点は，外部のLODとのリンクを介して関連情報を取得できる点にある．たとえば，上記のUCCコーヒー博物館のPOI情報からWikidata上のデータ

へのリンク^{★11}を辿ることで、英語、韓国、中国語での名称、施設の外観画像、公式サイトへのリンク、POIの分類（museum）などさまざまな情報を取得できる。

このような外部のLODから得られる情報は、より多くの観点からのデータ分析に活用できる。例として、LOD化したGPS移動軌跡を、外部のLODの1つであるWikidataの情報を利用して分析するSPARQLクエリを図7に示す^{★12}。図7(e)のクエリでは、Wikidata上で博物館（wd:Q33506というIDで定義されている）に分類されるPOIを訪問したユーザーの一覧を取得できる。このクエリ中で「SERVICE、<https://query.wikidata.org/sparql> {...}」と記述されている個所において、WikidataのSPARQLエンドポイント^{★13}から、クラス分類の情報を取得し、その結果を用いてGPS移動軌跡のLOD内を検索している。

```
PREFIX gtl-prop: <http://lodosaka.jp/iswc2016gtl-exp/prop#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
```

```
SELECT DISTINCT ?u where{
  {SELECT DISTINCT ?poi ?o where{
    ?s gtl-prop:poi ?poi .
    ?poi gtl-prop:source "Wikidata" .
    ?poi rdfs:seeAlso ?o.
    FILTER(regex(str(?o),"wikidata"))
  }
}
SERVICE <https://query.wikidata.org/sparql> {
  ?o wdt:P31 wd:Q33506 .
}
?s gtl-prop:poi ?poi;
  gtl-prop:user ?u.
}
```

(e) Wikidataで博物館クラス（wd:Q33506）に分類されているPOIを訪問したユーザーの一覧を取得するクエリ

```

PREFIX gtl-prop: <http://lodosaka.jp/iswc2016gtl-exp/prop#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>

SELECT DISTINCT ?cls ?label (count(DISTINCT ?u) AS ?c)
where{
  {SELECT DISTINCT ?poi ?o where{
    ?s gtl-prop:poi ?poi .
    ?poi gtl-prop:source "Wikidata" .
    ?poi rdfs:seeAlso ?o.
    FILTER(regex(str(?o),"wikidata"))
  }
}
SERVICE <https://query.wikidata.org/sparql> {
  ?o wdt:P31 ?cls .
  ?cls rdfs:label ?label.
  FILTER(lang(?label)="ja")
}
?s gtl-prop:poi ?poi;
  gtl-prop:user ?u.
}GROUP BY ?cls ?label
ORDER BY DESC(?c)

```

(f) 各POIのWikidataにおけるクラス分類ごとの、1回以上訪問したユーザー数をランキングするクエリ

図7 GPS移動軌跡のLODと外部のLOD (Wikidata) を連携させて分析するSPARQLクエリの例

さらに、図7(f)のクエリでは、同様な手法でWikidataでの各POIのクラス分類を日本語ラベルとともに取得し、クラス分類ごとに1回以上訪問したユーザー数のランキングが取得できる。このクエリの結果、訪問者数が3以上であったPOIのクラスを表4に示す。ホテル、鉄道駅といった国際会議に参加するために必ず訪問する必要がある場所に加え、博物館、講演、神社など、観光目的で訪問したと思われるPOIも見られる。

表4 POIのWikidataにおけるクラス分類ごとの訪問者数

	クラス ID	クラス名	訪問人数
1	wd:Q27686	ホテル	11
2	wd:Q55488	鉄道駅	11
3	wd:Q11303	超高層建築物	7
4	wd:Q33506	博物館	6
5	wd:Q12280	橋	5
6	wd:Q861951	警察署	5
7	wd:Q22698	公園	4
8	wd:Q23442	島	4
9	wd:Q845945	神社	3

このように外部LODの情報を利用することによって、さまざまな分析が可能になると期待される。

4.2 POIの情報源としてのオープンデータの有用性の分析

3.1節においてPOI情報の取得に用いた6種類のオープンデータについて、POIの情報源としての有用性という観点から分析する。第1章で述べたMICEの観点からGPSデータを分析するためには、適切なPOI情報を十分に収集することが重要となる。そこで、各オープンデータから取得したPOI情報が、GPSデータから滞在したPOIを判定する際に、どの程度利用されたかの分析を通して有用性を考察する。

表5に、3.3節で述べたGPSデータのLOD化の手法において、利用するオープンデータおよび滞在場所の判定に用いる距離の閾値dを変化させた際の判定結果の比較を示す。たとえば3.4節で述べたLOD化の条件（(1) - (6)の統合データを利用、d=100m）では、オープンデータから取得した730のPOI情報のうち146カ所を用いて、全GPSデータの82.88%の滞在場所が判定されたことが分かる。

表5 オープンデータから収集したPOI情報

	POI数	滞在場所の判定に利用されたPOI情報の数			滞在場所が判定できたGPSデータの数(割合)※		
		200	100	50	200	100	50
(1)観光施設情報	103	27	24	17	2,609 (64.84%)	1,585 (39.39%)	810 (20.13%)
(2)夜景スポット	22	13	12	9	218 (5.42%)	71 (1.76%)	39 (0.97%)
(3)ロケ地	88	42	36	29	3,380 (84.00%)	2,516 (62.52%)	1,434 (35.64%)
(4)野外彫刻	35	29	26	23	478 (11.88%)	201 (5.00%)	101 (2.51%)
(1)-(4)の統合データ	223	85	77	66	3,638 (90.41%)	2,759 (68.56%)	1,647 (40.93%)
(5)DBpedia Japanese	178	60	53	45	2,544 (63.22%)	691 (17.17%)	239 (5.94%)
(6)Wikidata	544	104	97	79	3,781 (93.96%)	2,697 (67.02%)	1,265 (31.44%)
(1)-(6)の統合データ	730	153	146	132	3,950 (98.16%)	3,335 (82.88%)	2,053 (51.02%)

※分析対処のGPSデータの総数は4,024であった。

利用したオープンデータごとの結果を比較すると、神戸市のオープンデータに加え、DBpedia JapaneseおよびWikidataを用いることで、滞在場所の判定ができるGPSデータの割合が10%前後向上していることが分かる。一方、その際に利用されたPOI情報の数は倍増しており、特にWikidataから取得できるPOI情報が多い。今回の実証実験では、ユーザの移動が神戸市内、特に、会議の会場と三ノ宮周辺に集中していた(4.1節参照)ため、神戸市のオープンデータが多くの範囲をカバーできていたが、より広範囲のPOI情報をカバーするためには、Wikidataの利用がきわめて効果的であると思われる。

5. おわりに

本稿ではGPSにより収集した移動履歴データを、オープンデータから得られたPOI情報と移動軌跡を表すRDFデータモデルに基づきLOD化する手法を提案した。国際会議ISWC2016において本手法の実証実験を実施し、十分な移動履歴データが収集できること、複数のオープンデータから得られたPOI情報を用いることで移動軌跡をLOD化し、容易に分析ができる形で公開できることを確認した。

この取り組みは、GPS移動履歴を例とした、自律測位によるセンシングデバイスから得られた情報を、オープンデータと組み合わせることにより分析する手法の実践と位置づけられる。また、POI情報を得るための情報源としてのオープンデータの有用性の比較は、複数のオープンデータを適切に統合する取り組みの実践例と言える。

今後、これらの実践例をさらに発展させることにより、多様な観点からのデータ分析の実現を目指す。具体的には、利用するオープンデータの情報をより詳細なものとすることや、センサの種類や変換方法の改良することにより、さまざまな展開が期待できる。

なお、本研究においては既存のオープンデータの活用の可能性を調べることを目的としており、GPSデータから取得したPOI情報が正しいか否かの検証を行っていない。たとえば、表4において「警察署」を訪問したユーザが5名いたという結果が出ているが、これだけの人数の参加者が実際に警察署を訪れたとは考えにくい。これは、POIの判定に滞在時間が考慮されていないため、単に近くを通過した場合であっても訪問したと判定されているためと思われる。またPOI判定に用いる距離の閾値dについても、現状では50～200mと大きめの値を設定しているが、より変換精度を上げるためには、POIの種類ごとにdを変化させるなどの手法の開発が必要と思われる。これらの変換精度を上げる手法の開発には、実際の訪問場所の被験者から聞き取るなど、データ収集の方法も合わせた検討が必要となる。この点については、今後の重要な課題の1つであると考えている。

謝辞 本研究の一部は、科学研究費補助金25280081, 17H01789 および16K12533の助成による。また実証実験に協力くださったISWC2016の関係者および、データ収集への協力加者の皆様に感謝します。

参考文献

- 1) 日本再興戦略—JAPAN is BACK—,
http://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/saikou_jpn.pdf
(2017年5月8日現在)
- 2) 観光立国推進閣僚会議, 観光立国実現に向けたアクション・プログラム,
<http://www.mlit.go.jp/common/001000830.pdf> (2013)
- 3) 観光立国推進基本計画,
<http://www.mlit.go.jp/common/000208713.pdf> (2012)
- 4) 国土交通省観光庁, 明日の日本を支える観光ビジョン,
http://www.mlit.go.jp/kankocho/topics01_000205.html (2017年5月8日現在)
- 5) 観光庁, 平成28年度MICEの経済波及効果及び市場調査事業報告書, <http://www.mlit.go.jp/common/001186651.pdf> (2017年8月7日現在)

- 6) 日本政府観光局 (JNTO) , 開催地をさがす,
https://mice.jnto.go.jp/place/unique_venue.php (2017年5月8日現在)
- 7) 観光ビッグデータを活用した観光振興について (中間とりまとめ) <https://www.mlit.go.jp/common/001086244.pdf> (2017年5月8日現在)
- 8) 国土交通省観光庁観光地域振興課, 平成27年度ICTを活用した訪日外国人観光動態調査事業実施報告書,
<http://www.mlit.go.jp/kankocho/shisaku/kankochi/gps.html> (2016)
- 9) Bulusu, N., Heidemann, J. and Estrin, D. : Gpsless Lowcost Outdoor Localization for very Small Devices, Personal Communications, IEEE, 7(5), (2000).
- 10) Anagnostopoulos, C.-N.E., Anagnostopoulos, I. E., Psoroulas, I. D., Loumos, V. and Kayafas, E. : License Plate Recognition From Still Images and Video Sequences : A Survey, IEEE Transactions on Intelligent Transportation Systems, Vol.9 No.3, pp.377-391 (2008).
- 11) Musa, A.B.M. and Eriksson, J. : Tracking Unmodified Smartphones Using Wi-Fi Monitors, In Proc. of SenSys '12 Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems, pp.281-294 (2012).
- 12) 総務省, 位置情報プライバシーレポート,
http://www.soumu.go.jp/main_content/000434727.pdf (2014)

脚注

- ☆1 <https://www.w3.org/RDF/>
- ☆2 <https://data.city.kobe.lg.jp/>
- ☆3 <http://ja.dbpedia.org/>
- ☆4 <https://www.wikidata.org/>
- ☆5 各データの位置情報に対し, 国土交通省が提供している「街区レベル位置参照情報 <http://nlftp.mlit.go.jp/isj/index.html>」を用いて住所を付与し, 住所が神戸市内となるものを抽出した.

☆6 IRIの省略表現のためPREFIXと呼ばれる接頭語が用いられる。たとえば, rdfs: という接頭語で

<http://www.w3.org/2000/01/rdf-schema#>を表すとしたとき, rdfs: <http://www.w3.org/2000/01/rdf-schema#label>と rdfs: labelは同じIRIを表す。

☆7 本データの作成時にPOI判定の元にしたオープンデータは日々更新されるため, それらのオープンデータを参照した日付によって, 生成されるデータ数は変化する。

☆8

<https://github.com/koujikozaiki/GPS2LOD/tree/GPS2LODforDP>

☆9 <http://lod.hozo.jp/repositories/GPS2LOD>

☆10 この結果を含めた可視化のサンプルは,

<https://github.com/koujikozaiki/GPS2LOD/tree/GPS2LODforDP>に公開している。

☆11 <https://www.wikidata.org/wiki/Q11251779>

☆12 これらのクエリは, 本研究で公開しているSPARQLエンドポイント<http://lod.hozo.jp/repositories/GPS2LOD>では, SERVICE構文の処理にエラーが発生して実行できなかった。そのため, Apache Jena

Fuseki<https://jena.apache.org/documentation/fuseki2/index.html>と別のRDFデータベースに本研究で構築, 公開したRDFデータを格納して動作を確認した。このように, SERVICE構文を用いて外部のLODと連携をするクエリは, 各LODが利用しているRDFデータベースの環境に依存して, 正しく動作しない場合があるので注意が必要である。

☆13 <https://query.wikidata.org/sparql>

古崎 晃司 (正会員) kozaki@ei.sanken.osaka-u.ac.jp

大阪大学産業科学研究所准教授, 博士(工学), オントロジー工学の基礎理論, オントロジー構築・利用環境の設計・開発, セマンティックWeb, Linked Data, 医療, 環境など各種領域におけるオントロジー開発・応用に関する研究に従事。

横山 輝明（非会員） ytel@kic.ac.jp

神戸情報大学院大学情報技術研究科講師. 奈良先端科学技術大学院大学情報科学研究科博士（工学）. サイバー関西プロジェクトメンバー. WIDEプロジェクトメンバー. 途上国におけるインターネット基盤運用技術や応用サービスの研究に関わる.

深見 嘉明（正会員） yoshiaki@rikkyo.ac.jp

立教大学大学院ビジネスデザイン研究科特任准教授. 博士（政策・メディア）. NPO法人Linked Open Data Initiative理事. 一般社団法人オープン&ビッグデータ活用・地方創生推進機構技術委員会委員, IPA共通語彙基盤ワーキンググループ委員.

投稿受付：2017年5月9日

採録決定：2017年9月4日

編集担当：定兼邦彦（東京大学）