

平成 27 年度修士論文

**放置自転車問題解決に向けた
循環型 LOD 構築システムの提案**

電気通信大学 大学院情報システム学研究科
社会知能情報学専攻

学籍番号 : 1451007
氏名 : 江上 周作

主任指導教員 : 川村 隆浩 客員准教授
指導教員 : 大須賀 昭彦 教授
指導教員 : 田原 康之 准教授

提出年月日 : 平成 28 年 1 月 28 日 (木)

概要

現在、国内では駐輪場施設の不足や問題意識の低さ、違法性の認識不足などのため放置自転車の発生が後を絶たず、地域問題・社会問題となっている。放置自転車は、街の美観を損なうだけでなく、歩行や車両通行の妨げ、交通事故、盗難の原因となっている。こうした放置自転車問題の解決に向けて、日々の放置自転車状況を Linked Open Data (LOD) として公開し、データ基盤を構築することが必要であると考え、この LOD を活用することで、自転車放置状況の可視化、最適な駐輪場の設置場所の提示、撤去活動の支援など、放置自転車問題解決に寄与するサービスの開発が可能になる。本研究では放置自転車問題解決に向けて必要なデータを収集し、LOD として統一化して公開し、さらに可視化することで市民の問題意識を向上させて次のデータ収集につなげる循環型システムを提案する。本研究ではまず、放置自転車問題に関する統一的な LOD スキーマ設計の方法論を示し、次に SNS から 813 件の実データと行政の Web サイトから放置自転車の台数に影響を与えるデータを収集した。設計した LOD スキーマに基づいて収集したデータを LOD 化した。さらに、データ収集の際に生じる欠損をベイジアンネットワークにより推定し、70.3% の精度で欠損値を推定した。推定結果を LOD に追加し、最終的に 219,804 トリプルの LOD として Web 上に公開した。最後に構築した LOD を可視化することで地域住民の問題意識向上と持続的なデータの収集につなげた。本システムにより放置自転車問題解決の一助となる有用なデータセットの構築が確認でき、他の地域課題・社会課題にも適応できる可能性を示した。

目次

第 1 章	はじめに	1
第 2 章	背景知識	4
2.1	オープンデータ	4
2.2	Linked Open Data	5
2.3	オントロジー	6
第 3 章	関連研究	8
第 4 章	循環型 LOD 構築システム	13
4.1	LOD スキーマ設計の方法論	14
4.1.1	ドメイン要件抽出	15
4.1.2	LOD スキーマの設計	16
4.2	一次データの収集	17
4.2.1	報酬型クラウドソーシングによるデータ収集実験	17
4.2.2	SNS を利用したデータ収集	20
4.3	スキーマ設計に基づく LOD の構築と公開	22
4.3.1	スキーマ設計に基づいた LOD のリアルタイム構築	22
4.3.2	放置自転車 LOD の公開	26
4.4	放置自転車台数推定による LOD 詳細化	28
4.4.1	欠損属性値の補完	30
4.4.2	ベイジアンネットワークによる台数推定	30
4.4.3	推定精度評価を通じたパラメータ設定	31
4.4.4	放置台数推定結果を用いた LOD の時系列補完	33
4.5	放置自転車 LOD の可視化	34

第 5 章	実験結果と考察	38
5.1	データ収集と LOD 構築の結果	38
5.2	ベイジアンネットワークによる台数推定の結果	40
第 6 章	まとめと今後の課題	43
	謝辞	45
	参考文献	46
	研究業績	50

目次

1.1	放置自転車問題解決に向けたエコシステムの提案	2
2.1	5-star Open Data の公開ステップ	6
3.1	The Linking Open Data cloud diagram (2014/08/30)	9
4.1	インスタンスを加えて設計した LOD スキーマ	16
4.2	作業予約・報告アプリケーションの画面例	18
4.3	ツイート Web アプリの動作画面	20
4.4	一次データ収集と LOD 化の手順	21
4.5	放置自転車 LOD 内の観測地点を検索する SPARQL クエリ	22
4.6	駐輪場の RDF グラフの例	24
4.7	観測地点と周辺施設の RDF グラフの例	25
4.8	気象情報の RDF グラフの一部	26
4.9	実際に構築される放置自転車 LOD の一部	27
4.10	ブラウザで表示される HR ページの例	29
4.11	観測地点と観測日時を全て取得する SPARQL クエリ	33
4.12	推定結果を基に構築される RDF (Turtle 形式)	34
4.13	可視化アプリケーションのスマートフォン画面例	36
4.14	欠損値推定以前以後の可視化アプリケーションの動作例	37
4.15	欠損値補完前後の可視化アプリケーションの表示例	37

表目次

4.1	抽出したキーワードのクラスタリング結果	14
4.2	ランサーズでの作業依頼内容	19
4.3	属性およびパラメータ修正前の Confusion Matrix	33
5.1	観測データの統計	39
5.2	10 回の 10 分割交差検証における適合率, 再現率, F 値の平均値	40
5.3	属性およびパラメータ修正後の Confusion Matrix (1)	42
5.4	属性およびパラメータ修正後の Confusion Matrix (2)	42
5.5	属性およびパラメータ修正後の Confusion Matrix (3)	42
5.6	属性およびパラメータ修正後の Confusion Matrix (4)	42
5.7	属性およびパラメータ修正後の Confusion Matrix (5)	42
5.8	属性およびパラメータ修正後の Confusion Matrix (6)	42
5.9	属性およびパラメータ修正後の Confusion Matrix (7)	42
5.10	属性およびパラメータ修正後の Confusion Matrix (8)	42
5.11	属性およびパラメータ修正後の Confusion Matrix (9)	42
5.12	属性およびパラメータ修正後の Confusion Matrix (10)	42

第1章

はじめに

現在、国内では日常的な社会課題の一つとして駅前の放置自転車が問題になっている。厚生労働省による「健康日本 21^{*1}」の取り組みや省エネルギー意識の高まりなどの影響もあり、我が国における自転車保有台数は昭和 45 年から平成 25 年までに 2.6 倍に増加している^{*2}。しかしながら、駐輪場施設の不足や問題意識の低さ、違法性の認識不足などのため放置自転車の発生が後を絶たず、歩行や車両通行の妨げ、盗難、街の美観を損なう原因となっている。

こうした放置自転車問題を解決するために、日々の放置自転車状況をオープンデータとして公開し、データ基盤を構築することが必要であると考え。また、データ形式の統一化やデータ同士の意味的な結合によりデータの利便性を向上させるため、オープンデータの完成形である Linked Open Data (LOD) が推奨されている。LOD とは、Resource Description Framework (RDF) で構造化され、二次利用可能な形式で Web 上に公開されたデータのことである。構造化されたデータ同士をリンクすることで、「データの Web」を構築することが目標とされている。このオープンデータおよび LOD の二次利用により、例えば、自転車放置状況の可視化、最適な駐輪場の設置場所の提示、撤去活動の支援などが可能になる。しかし、現時点では放置自転車のデータは広範囲かつ年単位のものしか公開されておらず、サービスが放置自転車のデータを知識として活用することが難しい状態である。また、その他の公開されているデータも形式が多種多様であり、データ形式の統一化やスキーマ（データ構造）の定義はオープンデータの再利用性向上のための重要課題となっている。

^{*1} <http://www.kenkounippon21.gr.jp/>

^{*2} <http://www.soumu.go.jp/menunews/s-news/94984.html>

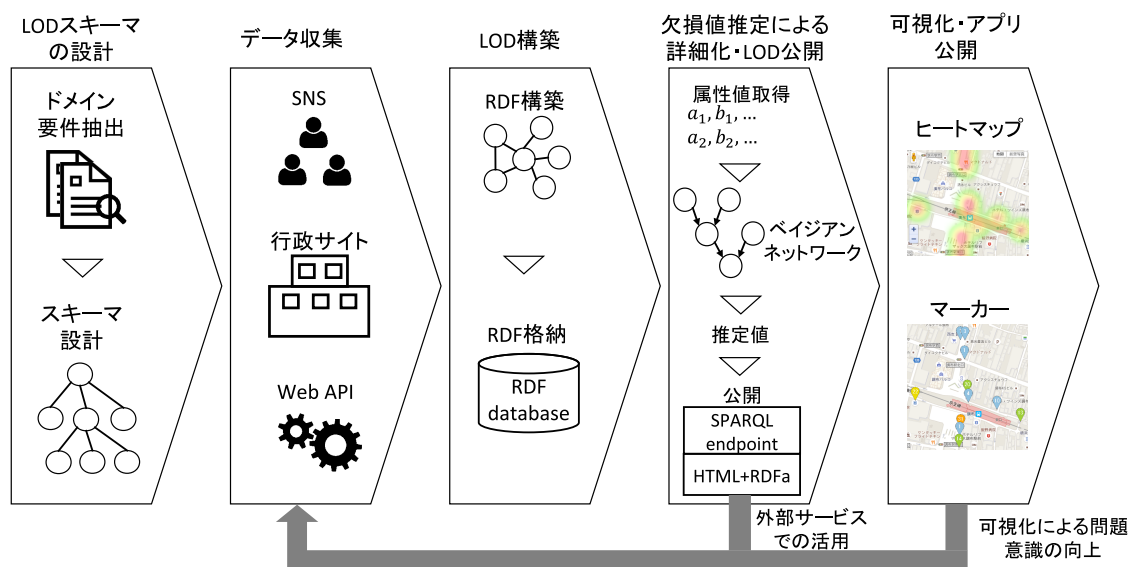


図 1.1 放置自転車問題解決に向けたエコシステムの提案

形式に関しては、海外では異なる市のオープンデータを Linked Data^{*3}として統合し、その際に生じる欠損値を補完して再配布する研究がされている [1]。これは、再利用性に欠けたオープンデータの利活用を促進させる有用な手法であると考えられる。一方、オープンデータを活用して放置自転車問題解決に寄与するサービスを開発するためには、既存のオープンデータに加えて未だデータ化されていない時間的、または空間的により詳細なデータも必要である。そこで、本研究ではまず放置自転車問題に関する統一的な LOD スキーマ設計の方法論を示し、次に SNS や行政の Web サイトから実際のデータを収集した上で、欠損値を推定して LOD として統合し、可視化する枠組を提案する (図 1.1)。欠損値補完により時系列的に疎なデータを密なデータにすることで、視覚的に滑らかで動的な可視化が可能になる。これによって社会的な問題意識が高まれば、より多くのデータが集まるようになり、LOD が持続的に発展していくと同時に、放置自転車問題解決の一助になることが期待できる。

したがって、本論文の貢献は、放置自転車のように日常的に生じる地域課題解決に向けて必要なデータを収集し、統一された形式で公開し、発展させることを目的としたシステムの提案と事例紹介である。具体的には、

1. Web 情報から LOD を構築する方法論の提案

^{*3} <https://www.w3.org/DesignIssues/LinkedData.html>

2. 行政, SNS からの放置自転車 LOD の構築と公開
3. 欠損属性値の補完, 推定手法の開発と評価
4. 放置自転車問題可視化アプリケーションの開発と評価

の4点にある。しかし、こうした施策により実際に放置自転車への意識が高まり、問題解決に繋がるかどうかには長期間に渡る社会実験が必要であり、現時点ではまだ検証できていない。

以下、本論文では、まず2章で背景知識を説明し、3章で関連研究について述べる。4章で提案システムについて述べ、5章で実験結果と考察について述べる。最後に6でまとめと今後の展望を述べる。

第2章

背景知識

本章では本研究における基礎技術および背景知識について述べる。

2.1 オープンデータ

オープンデータの普及を行う非営利団体、Open Knowledge Foundation^{*1}作成のオープンデータハンドブック^{*2}によると、オープンデータとは Open Definition^{*3}で定義されるオープン性を満たすデータとされている。オープンデータハンドブックによると、最も重要な要件は以下の三点とされている。

1. 利用できる、そしてアクセスできる
2. 再利用と再配布ができる
3. 誰でも使える

これらの要件を満たしたオープンデータにより、データの相互運用性が高まり新たな価値が創出される。オープンデータはこれらの要件を満たすために、Creative Commons^{*4}のライセンスを付与するケースが多い。特に、クレジット表示をすれば商用・非商用を問わず改変、再利用、再配布が可能な CC-BY^{*5}や再配布時に同じライセンスを継承する CC-BY-SA^{*6}などが推奨されている。

*1 <http://okfn.org/>

*2 <http://opendatahandbook.org/guide/ja/what-is-open-data/>

*3 <http://opendefinition.org/>

*4 <https://creativecommons.org/>

*5 <http://creativecommons.org/licenses/by/4.0/>

*6 <http://creativecommons.org/licenses/by-sa/4.0/>

2.2 Linked Open Data

Linked Open Data (LOD) とはデータを機械可読形式で Web 上に公開し、データ同士をリンクさせることで「データの Web」を構築する取り組み、およびそれらのデータを指す。LOD はオープンデータかつ Linked Data と定義することができる。Linked Data とは、ティム・バーナーズ＝リーの定義した以下の基本原則^{*7}を満たすデータである。

1. あらゆる事物に URI を付与すること
2. 誰でも事物の内容が確認できるように、URI は HTTP 経由で参照できること
3. URI を参照した時は、標準の技術 (RDF や SPARQL 等) を使用して関係する有用な情報を利用できるようにすること
4. より多くの事物を発見できるように、他の URI へのリンクを含めること

RDF (Resource Description Framework) とは W3C (World Wide Web Consortium) の推奨するデータ標準化のための枠組みである。RDF ではデータを主語・述語・目的語の三つ組 (トリプル) で表現し、各リソースは URI (Uniform Resource Identifier) の形式で記述される。また、N-Triples, RDF/XML, Turtle などの様々なデータシリアライゼーションフォーマットが存在する。Linked Data において RDF がデータ記述のデファクトスタンダードとなっており、RDF で記述されたデータ同士がリンクすることで Linked Data が形成される。SPARQL (SPARQL Protocol And RDF Query Language) は RDF に対して問い合わせが可能なクエリ言語である。また、ティム・バーナーズ＝リーはオープンデータの完成度の基準として図 2.1 に示す 5-star Open Data^{*8}を提案している。この基準では、一つ星オープンデータとしてはデータのオープン性のみが要求され、構造化の有無は問わず PDF や画像なども可としている。二つ星オープンデータではデータの構造化が要求される。三つ星オープンデータでは独自仕様でない形式が要求され、例として CSV (Comma-Separated Values) が挙げられる。四つ星オープンデータでは物事の記述に URI を使用することが要求される。これにより他のデータとリンクすることが可能になる。五つ星オープンデータでは他のデータへリンクしていることが要求される。この基準における五つ星オープンデータが LOD であり、オープンデータの完成形と言える。すなわち、Open Definition で定義されるオープン性を満たした状態で公開される Linked Data が LOD である。

^{*7} <http://www.w3.org/DesignIssues/LinkedData.html>

^{*8} <http://5stardata.info>

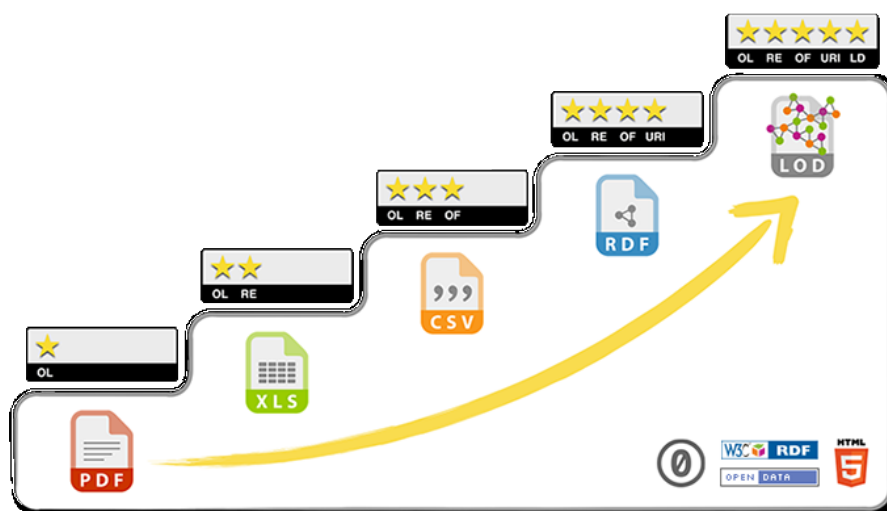


図 2.1 5-star Open Data の公開ステップ

2.3 オントロジー

溝口は、オントロジーを「人間が対象世界をどのように見ているかという根源的な問題意識をもって物事をその成り立ちから解きあかし、それをコンピュータと人間が理解を共有できるように書き記したもの」と定義している [2]。人工知能の立場からは「概念化の明示的な規約」と定義されている^{*9}[2]。RDF がコンテンツのメタデータ記述に利用されるのに対し、オントロジーはそのメタデータの意味を定義する概念体系・知識体系として用いられることが多い。オントロジーは代表的に、上位オントロジー、ドメインオントロジー、タスクオントロジーに分類できる。上位オントロジーはオントロジーの中でも抽象度が高く、全てのドメインの上位に位置するオントロジーである。例として、DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) [3] や YAMATO (Yet Another More Advanced Top-level Ontology) [4] などが挙げられる。ドメインオントロジーは上位オントロジーの下位にあたり、対象ドメインのオントロジーである。例として、Gene Ontology[5] や医療オントロジー [6] などが挙げられる。タスクオントロジーはタスクの実行過程の概念を構造化したオントロジーである。例として故障オントロジー [7] が挙げられる。

^{*9} <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>

また、W3CはWebオントロジー言語としてOWL (Web Ontology Language)^{*10}を規格化しており、セマンティック Web や LOD の分野において主に使用される。OWLは記述論理 (Description Logic) を基に設計されており、OWLによりオントロジーを記述することで記述論理に基づいた推論が可能になる。記述論理は知識表現のための論理体系である。

^{*10} <https://www.w3.org/TR/owl-overview/>

第3章

関連研究

近年、LODを構築・活用する動きが活発化しており、図3.1に示す Linking Open Data cloud diagram^{*1}から分かるように、様々な分野のデータがLOD化されている。例えば、WikipediaデータのLODであるDBpediaを中心に、英国政府統計情報^{*2}、地理情報分野のLinkedGeoData[8]、ライフサイエンス分野のBio2RDF[9][10]など様々な分野でLODが構築され、これらを活用した研究やサービスの開発などが進んでいる。国内でもLODの取り組みは活発化しており、日本語WikipediaのLODであるDBpedia Japanese[11]を中心として、博物館情報のLODであるLODAC Museum[12]、横浜市内の地域・芸術分野のLODであるヨコハマ・アート・LOD^{*3}、締結部品のLODであるねじLOD[13]など、様々なLODが公開されている。しかし、日常的な地域課題についてはLOD化が進んでおらず、本研究で構築する放置自転車LODがこの分野における最初の時系列LODである。

LODの基となるデータ収集手法としては、クラウドソーシングやGame With A Purpose(GWAP)[14]の活用が挙げられる。クラウドソーシングとは、インターネットを通して不特定多数の人(Crowd)に作業を依頼(Sourcing)する仕組みであり、近年世界的に普及している。クラウドソーシングには金銭を支払う報酬型と、ボランティアによる無報酬型クラウドソーシングがある。前者の代表的な例として米国のAmazon MTurk^{*4}が挙げられる。後者の代表的な例としてはWikipedia^{*5}が挙げられる。GWAPとは、プレイヤーがゲームをプレイすることで、その副産物として有益な資源が生成されるような

^{*1} <http://lod-cloud.net>

^{*2} <https://datahub.io/dataset/statistics-data-gov-uk>

^{*3} <http://yan.yafjp.org/lod>

^{*4} <https://www.mturk.com/mturk/welcome>

^{*5} https://en.wikipedia.org/wiki/Main_Page

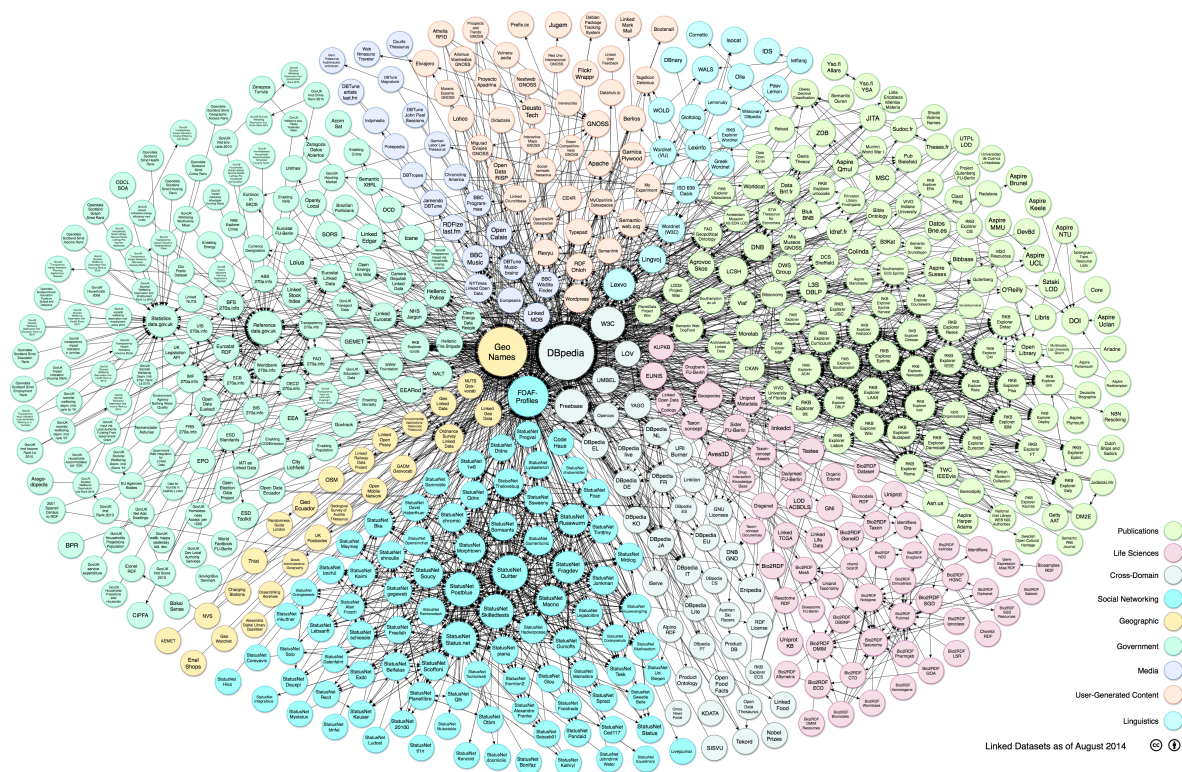


図 3.1 The Linking Open Data cloud diagram (2014/08/30)

ゲームである。OpenStreetMap^{*6}はクラウドソーシングによりオープンな地図を作成するプロジェクトである。誰でも編集が可能であり、作成されたデータはオープンデータとして公開されている。FixMyStreet^{*7}は道路状況や不法投棄などの地域の問題を報告するプラットフォームである。Web上に存在していない地域問題の情報をクラウドソーシングにより収集し可視化することで、行政だけでなく広く一般に問題が共有され、従来よりも地域問題の早期解決が見込まれる。国内でもFixMyStreet Japan^{*8}が開発されており、誰でも地域問題を報告することが可能になっている。Zookら [15] は、2010年のハイチ地震直後に公開された衛生画像のオープンデータを、クラウドソーシングによりOpenStreetMapとマッシュアップした事例について報告している。これにより救援活動に役立つ地図が作成され、新たに作成されたデータはオープンデータとして公開された。この事例は、オープンデータにより遠隔地の人でも災害の救援活動に加担できること

*6 <https://www.openstreetmap.org>

*7 <http://fixmystreet.org/>

*8 <https://www.fixmystreet.jp/>

を示し、オープンデータの社会的意義が顕著に現れることとなった。鈴木ら [16] は屋内 Location Based Service(LBS) の実現のために、オープンな屋内地図情報を構築する研究を行っている。この研究ではフロアマップを基にクラウドソーシングのアプローチで屋内構造地図の収集を目指している。東田ら [17] は、金銭的な報酬を支払うタイプのクラウドソーシングにより、カーナビゲーションや道案内サービスの主要コンテンツになる POI を収集している。調査地点を無作為に 400~4000 件選定し、ワーカーに実際に現地に出向いてもらいデータを収集する。Celino ら [18][19] は、GWAP と Human Computation [20] により、Linked Data を検証・修正・追加する手法を提案している。Human Computation とはコンピュータで処理することが困難な作業を人手で処理する技術であり、クラウドソーシングや GWAP などと組み合わせて使用されることが多い。[18] では写真選択クイズ形式のスマートフォンゲームを開発しており、ゲームをプレイすることによりバックグラウンドで LOD の検証やデータの追加が行われる。[19] ではゲーム性を更に高めて、Urbanopoly*⁹ というモノポリ形式のゲームを開発している。このゲームでも同様に、バックグラウンドで LOD の検証やデータの追加を行う。Urbanopoly は Google Play ストア*¹⁰ 上で公開されており、誰でもプレイすることが可能になっている。しかし、本研究で対象とする社会課題に関するデータは地理的に分散した時系列データであることもあり、これらのクラウドソーシングや GWAP だけでは継続的にデータを集めることが難しい。本研究においても当初、クラウドソーシングにより放置自転車情報を収集することを試みた。クラウドソーシングサービス「ランサーズ*¹¹」を使用し、観測地点と一定の期間を定めた上で、ワーカーが自由に作業できるようにマイクロタスク型クラウドソーシングを実施した。しかしながら、調布駅、吉祥寺駅で計 4 回、段階的に報酬価格を上げてクラウドソーシングを実施し、最終的に 1 件の報告につき 300 円まで報酬価格を上げたものの、放置自転車のデータを得ることはできなかった。Kaufmann ら [21] は、Amazon MTurk のワーカーに対して調査を行った結果、ワーカーのモチベーションは金銭的な報酬が最大であったと述べており、本ケースにおいては報酬と現地調査という手間のバランスが合わなかった可能性がある。我々はこの経験から、持続的なデータ収集における費用対効果を考慮し、場所が限定された上で外出を伴う作業には、クラウドソーシングサービスを用いたマイクロタスク型クラウドソーシングは不向きであると判断した。そのため、本研究では、まず SNS、行政の Web サイトからデータを収集し、ベイジアンネットワークにより欠損値を推定しながら時系列 LOD を構築した。

*⁹ <https://play.google.com/store/apps/details?id=it.cefriel.swa.urbanopoly.client.game>

*¹⁰ <https://play.google.com/store>

*¹¹ <http://www.lancers.jp/>

また、データの収集手法として実センサーを用いる研究もあり、海外ではセンサデータやストリームデータなどの時系列データを Linked Data 化する動きが活発化している。Patni ら [22] はセンサデータを RDF に変換するフレームワークを提案している。Atemezing ら [23] はスペインの気象庁から得られるセンサデータを Linked Data 化している。国内では、ストリーム推論により衝突可能性を検知する先進運転支援システム [24][25][26] の研究が行われている。この研究では、自動車に搭載されたセンサから得られるデータを RDF ストリームとして取得し、設計したオントロジーと SWRL^{*12}規則推論器、C-SPARQL[27] を用いて知的意思決定を行うシステムを提案している。Lopez ら [28] は静的データおよびセンサデータを Linked Data で管理・提供するプラットフォーム「QuerioCity」を提案している。この研究では、センサーからストリームデータを取得し、IBM InfoSphere Streams^{*13}と C-SPARQL を使用することでリアルタイムに RDF を提供している。実際に IBM のサポートにより、アイルランドのダブリン市のデータポータル「Dublinked^{*14}」で運用されており、バスの路線、位置、遅延、混雑などの情報を 20 秒毎に更新して提供している。しかし、この手法では実センサを使用しておりコストがかかるため、本研究のように草の根運動的に収集が必要なデータに関しては、同様の取り組みを行うことは難しい。

また、本研究と同様に社会データを補完・推定する取り組みとしては、前述した [1] が挙げられる。Bischof らは市のオープンデータを収集し、それらを Linked Data として統合し再配布する仕組みを提案している。この研究では、DBpedia[29]、Eurostat^{*15}、United Nations Statistics Division (UNSD)^{*16}などの複数のデータソースからデータを収集し、形式の異なるこれらのデータをオントロジーに基いて Linked Data 化する。その際に生じる欠損値を統計的回帰法と主成分分析を組み合わせた手法により推定している。さらに、異なるデータソースから欠損値を推定する Cross Data Set Prediction を提案している。一次データは、複数の異なるデータソースから定期的にクロールすることで収集している。データの詳細さについては重点を置いておらず、主に統計情報などの市のオープンデータを収集対象としている。しかし、本研究のように、詳細度の高い時系列データが必要な場合においては、定期的なクロールのみで解決することは難しい。また、Bischof らの欠損値推定手法は、同種のデータセットが異なるデータソースに複数存在する場合に

*12 <https://www.w3.org/Submission/SWRL/>

*13 <http://www-03.ibm.com/software/products/ja/ibm-streams>

*14 <http://www.dublinked.ie/>

*15 <http://ec.europa.eu/eurostat>

*16 <http://unstats.un.org/unsd/default.htm>

有用であるが，本研究のように同種のデータセットが他に存在しない場合には適用が難しい．そこで，本研究では SNS から位置情報と日時情報を含む放置自転車のデータをリアルタイムに収集し，時系列的に疎な部分を，定義した LOD スキーマに基づくベイジアンネットワークにより推定・補完することを試みた．

第 4 章

循環型 LOD 構築システム

本研究では、まず図 1.1 に示すように放置自転車の現状を表す LOD（以下、放置自転車 LOD）を構築するため、必要な要件を抽出した上でスキーマを設計する。LOD の構築において、実用面を重視するとセマンティクスの付与や意味的整合性が疎かになることがあり、オントロジーに基づく推論や意味的検索が困難になる。一方で、意味的整合性や知識表現を重視すると、構造が複雑になり実用性が低くなることがある。そこで、本研究ではオントロジーのセマンティクスに基づいて意味的整合性を確保しつつ、放置自転車問題解決という実用面を考慮して設計する。この手順を日常的に生じる地域課題に関する LOD スキーマ設計の方法論として示す。次に、SNS や行政の Web サイトから放置自転車に関連するデータを収集する。データ形式の異なるこれらのデータを設計したスキーマに基いて LOD 化する。しかし、データの収集にはコストの問題から限界があり、必ず時間的または地理的な欠損が生じる。信頼性が高く、サービスでの活用の際に十分な量のデータを得るためには、この欠損データ（放置自転車の未観測台数）を高い精度で推測する必要がある。この欠損データは放置自転車と関係する様々な属性値を基に推測することができると考えられるが、この属性値をどのように決定するか、また、この属性値にも欠損がある場合にどのように対処するかといった課題が存在する。そこで、放置自転車の要因を LOD スキーマのセマンティクスに基づいて選択し、それらを属性値としたベイジアンネットワークにより欠損データを推定する。これにより、欠損を補う形で放置自転車に関する時系列 LOD を構築する。さらに、構築した放置自転車 LOD を用いて、自転車の放置状況を可視化するサービスを開発する。この可視化サービスにより、地域住民に放置自転車の問題意識を喚起し、放置自転車に関する更なる情報提供を促す。すなわち、本研究は以下の 5 つのステップに分けられ、2 から 5 までを繰り返し実行する。

表 4.1 抽出したキーワードのクラスタリング結果

プロパティ	キーワードのカテゴリ	キーワード
place, factor	周辺施設	商店街, 大規模店舗, 官公署, 学校, 図書館, 公会堂等公的施設, 百貨店, スーパーマーケット, 銀行, 遊技場, パチンコ屋, 複合ビル
	駐輪場	利便性, 安全性, 快適性, 供給実体, 料金, 駅までの距離, 附置義務駐車場
	鉄道	鉄道, 自宅から駅までの距離
	保管場所	容量
time, factor	時間	平日, 休日, 時間帯
product	事故	交通渋滞, 車両通行の妨げ
factor	目的	通勤, 通学, 最終目的地
	天気	雨天時の代替交通手段の有無
	自転車	低価格化, 質の低下, 維持費, 種類

1. LOD スキーマの設計
2. 観測データと属性データの収集
3. スキーマ設計に基づいた LOD の構築
4. ベイジアンネットワークによる放置台数の予測・推定
5. LOD の可視化

このように、データを収集し欠損を補いながら LOD を構築し、それを可視化することで地域住民に対して放置自転車問題への意識を喚起し、新たなデータの収集につなげる。このように放置自転車問題に関する LOD を持続的に構築し、発展させるエコシステムを形成する。

4.1 LOD スキーマ設計の方法論

オントロジー研究においては、オントロジー構築の方法論が長く議論されている。我々はタスクオントロジーとドメインオントロジー構築の方法論である Activity-First Method[30][31] を参考に、よりドメイン要件とサービスでの活用に焦点を当てた LOD ス

スキーマ設計の方法論を提案する。この方法論に基づいて放置自転車問題のスキーマを設計し LOD を構築する。本研究で提案する LOD スキーマ設計の方法論は以下の 2 つのステップから構成されている。

1. ドメイン要件抽出
 - a. 対象とする地域課題をモデル化する既存オントロジーを選択
 - b. 検索エンジンを用いた記事検索
 - c. 記事からキーワードを抽出
 - d. キーワードのクラスタリング
2. スキーマ設計
 - a. 既存オントロジーを軸としてスキーマを設計
 - b. クラスタリング結果を基にインスタンスとプロパティと値の追加

4.1.1 ドメイン要件抽出

まず、対象とする地域課題を LOD 化する上で骨組みとなる既存オントロジーを選択する。LOD をオントロジーに基づいて構築することで、オントロジーのセマンティクスを用いた推論が可能になることに加えて、LOD ごとに異なるデータ構造を理解する手間が省けるという利点がある。これは LOD をデータセットや API として活用する際に利便性を向上させる。放置自転車問題解決を目的とした時に、オントロジーに基づいた LOD の意味的整合性だけでなく、データの実用性を高める必要がある。そこで、本研究では放置自転車をイベントとして捉えて、構造的に実用性の高い Event Ontology^{*1} (以下, EO) を選択した。

次に、地域課題名を検索キーワードにして、検索エンジンで Web 上の記事を検索する。検索結果上位 10 件の記事と各記事の参考文献を調査し、選択したオントロジーのプロパティ部分に相当するキーワードを抽出する。本研究では“放置自転車”のワードで Google から記事を検索し、その結果の記事および各記事の参考文献から、放置自転車問題において EO のプロパティである, location, time, active, agent, factor, product について記述している文からキーワードを抽出した。

抽出したキーワードを表 4.1 のように手動でクラスタリングし、次のスキーマ設計に用いる。これらのキーワードが放置自転車問題を形作る要件となる。

^{*1} <http://purl.org/NET/c4dm/event.owl>

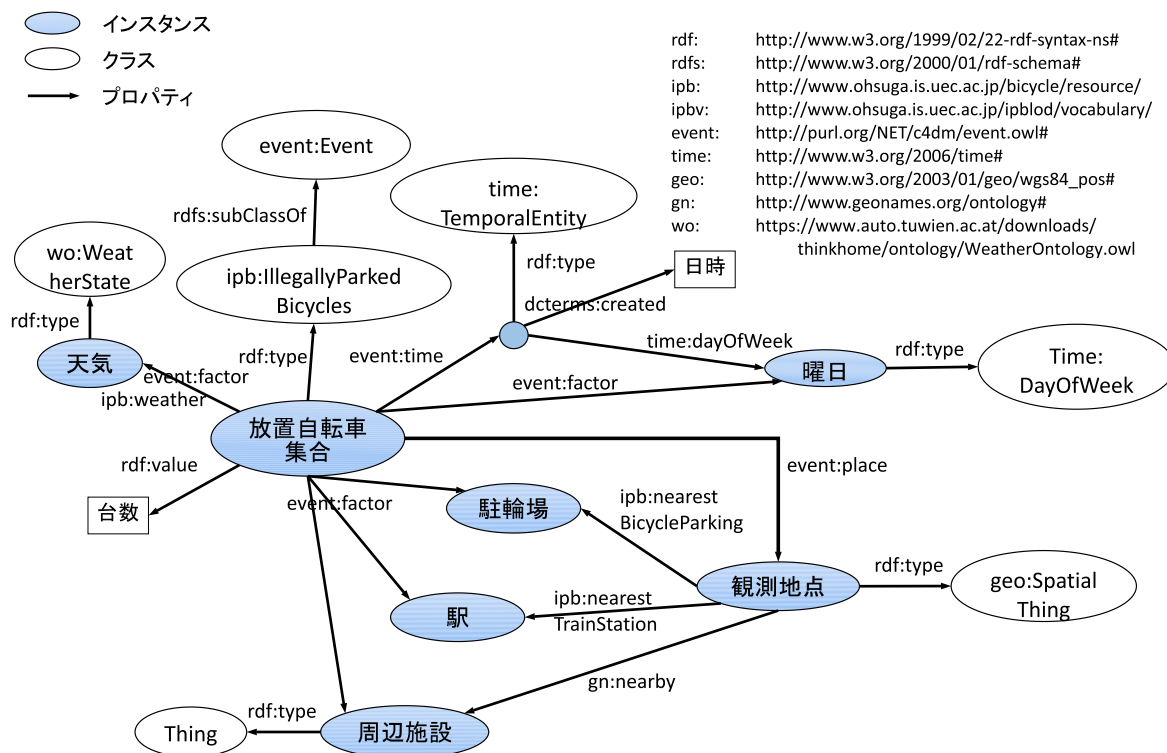


図 4.1 インスタンスを加えて設計した LOD スキーマ

4.1.2 LOD スキーマの設計

前のステップで選択した既存のオントロジーを軸として、LOD のクラスを設計する。本研究では EO のモデルに基づいて放置自転車 LOD のクラスを設計する。放置自転車 LOD の要件を以下に記述論理で示す。

$$IllegallyParkedBicycles \sqsubseteq Event \quad (4.1)$$

$$IllegallyParkedBicycles \sqsubseteq \exists place.SpatialThing \quad (4.2)$$

$$IllegallyParkedBicycles \sqsubseteq \exists time.TemporalEntity \quad (4.3)$$

$$IllegallyParkedBicycles \sqsubseteq \exists weather.WeatherState \quad (4.4)$$

$$IllegallyParkedBicycles \sqsubseteq \exists agent.Agent \quad (4.5)$$

$$IllegallyParkedBicycles \sqsubseteq \exists factor.Thing \quad (4.6)$$

$$IllegallyParkedBicycles \sqsubseteq \exists product.Thing \quad (4.7)$$

$$IllegallyParkedBicycles \sqsubseteq \exists value.Integer \quad (4.8)$$

IllegallyParkedBicycles クラスは放置自転車の集合を意味し、Event は Event クラスを意味する。式 4.1 では IllegallyParkedBicycles クラスは Event クラスのサブクラスであることを意味する。式 4.2 では IllegallyParkedBicycles クラスのインスタンスは必ず place (場所) として SpatialThing クラスのインスタンスを持つことを意味しており、以降の式で time (時間), weather (天気), agent (人物), factor (要因), product (引き起こす事象), value (放置自転車の台数) も同様に定義している。

次に表 4.1 を参考に LOD のインスタンスを設計する。4.1 にインスタンスを加えて設計した放置自転車 LOD のスキーマの全体図を示す。4.1 のキーワードのカテゴリが場所、時間、天気などのインスタンスに相当し、IllegallyParkedBicycles クラスのインスタンスから EO のプロパティで意味付けされている。また、インスタンス同士の間接関係を独自に定義したプロパティで意味付けしている。実際に LOD 化するには、4.1 のインスタンスには実データが入ることになる。但し、ここでは自転車を放置した人物をデータとして取得することはできないため、event:agent クラスは図から省略している。また、各観測地点における放置自転車が引き起こす事象も取得することが困難であるため、event:product クラスも除いている。同様に、保管場所の容量、放置者の目的、自転車の価格についても除外している。これは放置自転車 LOD スキーマとしては存在しているが、実際にはデータ取得が困難であるため本論文では除外している。このように、既存オントロジーを基に LOD の根幹部分を設計した後に、ドメイン要件として抽出したキーワードを追加していくことで全体の LOD スキーマを設計した。

4.2 一次データの収集

4.2.1 報酬型クラウドソーシングによるデータ収集実験

本研究では、まず初めに報酬型クラウドソーシングによる放置自転車データ収集実験を行った。クラウドソーシングサービスとしてランサーズ^{*2}を利用した。ランサーズでは、プロジェクト方式、コンペティション方式、タスク方式の3つの方式で作業を依頼することができる。プロジェクト型は納期や報酬を定めて作業を依頼し、応募された人数の中からプロフィールや提案内容を参考に作業員（ワーカ）を一人選ぶ方式である。依頼主は固定報酬と時間報酬を選択することができる。プロジェクト方式の例としてシステムの開発業務やアプリケーション開発業務が挙げられる。コンペティション方式は依頼内容に対して複数のワーカが成果を投稿し、その中から報酬を支払うワーカを依頼主が決定する方式

^{*2} <http://www.lancers.jp/>



図 4.2 作業予約・報告アプリケーションの画面例

表 4.2 ランサーズでの作業依頼内容

	1回目	2回目	3回目	4回目
対象地域	調布駅	吉祥寺駅	吉祥寺駅	吉祥寺駅
観測地点数	2	5	3	3
募集期間	05/05-05/07	05/26-06/03	06/05-06/08	06/11-06/14
報酬 (円/件)	50	100	250	300
調査時間帯	朝 (08:00-10:00)			
	昼 (14:00-16:00)			
	夜 (20:00-22:00)			

である。コンペティション方式の例として企業や団体のロゴ作成や看板のデザイン作成などが挙げられる。タスク方式は一つの依頼内容に対して複数のワーカが同時に作業し、各ワーカに対して報酬を支払う形式である。タスク型の例としてアンケートの解答やデータ入力などが挙げられる。本研究では放置自転車の台数報告をタスク方式で依頼し、東京都調布駅付近と東京都吉祥寺駅付近の放置自転車データの取得を試みた。作業概要を表 4.2 に示す。データの偏りを避けるため、各観測地点における各観測時間帯に作業できる人数は 3 人までとし、事前に作業の予約ができる図 4.2 のアプリケーションを開発した。このアプリケーションでは、作業内容確認のためユーザ管理を行い、新規ユーザ登録・ログイン後に作業予約が可能になる。ログイン後に観測地点のマークをクリックするとスケジュール表が表示され、予約人数が 2 人以下の日時に作業予約することが可能である。作業予約済みの場所では台数報告が可能となり、観測日当日に位置情報を付けて放置自転車の台数を報告する。位置情報、日時情報が既定のものと大幅に異なる場合は送信不可能とすることで、不正な作業をフィルタリングしている。データ送信後に放置自転車の台数、観測日時、位置情報、ユーザ情報が、認証番号 (6 桁の乱数) とともにサーバのデータベースに保存される。マイページでは作業予約状況の確認をすることができ、作業報告後に認証番号が表示される。この認証番号をランサーズの作業ページから報告することで作業の正当性を確認して報酬を支払うことができる。このように、データの偏りと不正作業を防ぐために、開発したアプリケーションを使用してデータ収集を行った。

しかしながら、表 4.2 のように段階的に報酬額を上げて作業を依頼したところ、作業の予約はあったものの実際にデータを取得することはできなかった。この原因として、作業に対して報酬額が合っていなかった可能性がある。また、クラウドソーシングサービスで



図 4.3 ツイート Web アプリの動作画面

は主に在宅作業が中心であり，本ケースのように外出を伴う作業はランサーズのユーザーの性質と合っていなかったことが考えられる．この経験から，本実験のように地域を限定した上で外出を伴う作業に報酬型クラウドソーシングは適さないと判断した．そこで，放置自転車に問題意識を持っている不特定多数の人々の作業を期待し，SNS を利用したデータ収集を検討した．

4.2.2 SNS を利用したデータ収取

4.2.1 の結果を受けて，SNS を利用したデータ収集を行った．具体的には，Twitter 上で位置情報，写真，ハッシュタグを付けた放置自転車台数報告のツイートを募集し，放置自転車のデータを収集した．しかし，携帯電話やスマートフォンは比較的安価な GPS チップを搭載しているため，天候や GPS 干渉地域などの要因により精度に問題が発生する [32]．そこで，位置情報を取得後に修正してツイートすることが可能な Web アプリケーション^{*3}を作成，公開し，このアプリケーションを使用してツイートしてもらうように呼びかけた．4.3 に作成したアプリケーションの動作例を示す．Twitter アカウントを使ってログイン後に 4.3 の画面に遷移し，位置情報追加ボタンを押すと地図上の現在位置にマークが表示される．このマークをドラッグすることで位置情報の修正が可能であるた

^{*3} http://www.ohsuga.is.uec.ac.jp/bicycle/tweet_submit

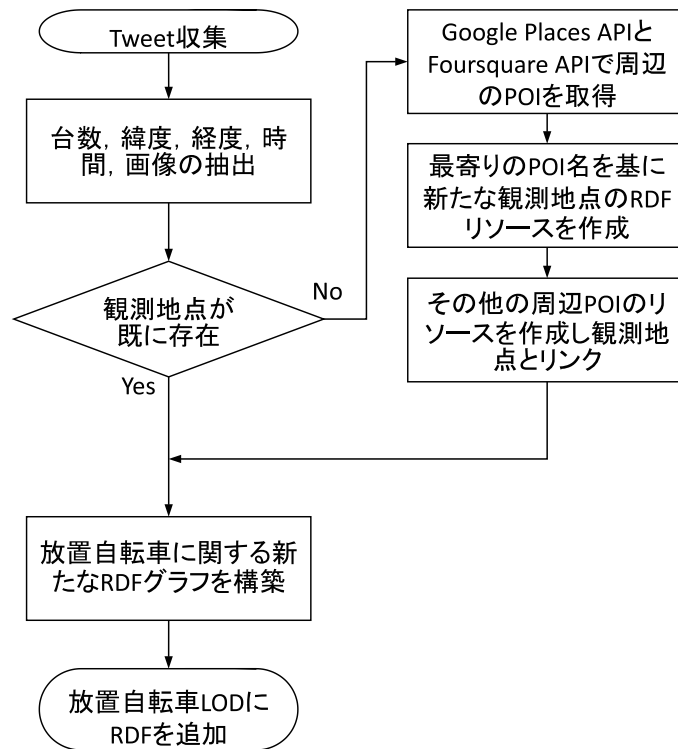


図 4.4 一次データ収集と LOD 化の手順

め、正確な位置情報を送信することが可能である。位置情報を追加し、放置自転車の台数を入力し、カメラを起動させ写真を撮った状態で送信すると、Twitter に位置情報、写真、ハッシュタグを付けた放置自転車台数報告のツイートが投稿される。実際に得られたツイート数は 813 件であった。

さらに、4.1.1 で要件として得られたキーワードに関して、周辺施設の情報を Google Places API^{*4}と Foursquare API^{*5}から取得し、駐輪場の情報を市区町村の Web サイトから収集する。気象情報は気象庁の Web サイト^{*6}から収集した。

```

PREFIX rdf: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
PREFIX ngeo: <http://geovocab.org/geometry#>
PREFIX ogcgs: <http://www.opengis.net/ont/geosparql#>
SELECT ?s ?lat ?lng FROM <http://www.ohsuga.is.uec.ac.jp/ipblod>
WHERE {
    ?s rdf:type geo:SpatialThing .
    ?s ngeo:geometry [
        ogcgs:asWKT ?g
    ];
    geo:lat ?lat;
    geo:long ?lng .
    FILTER(<bif:st_intersects> (?g, <bif:st_point> (経度, 緯度),半径))
}

```

図 4.5 放置自転車 LOD 内の観測地点を検索する SPARQL クエリ

4.3 スキーマ設計に基づく LOD の構築と公開

4.1 で述べた LOD スキーマに従い、4.2 で収集した放置自転車に関する一次データを LOD 化する。

4.3.1 スキーマ設計に基づいた LOD のリアルタイム構築

LOD 化の手順を 4.4 に示す。まず、サーバプログラムが特定のハッシュタグと位置情報のついた放置自転車の報告ツイートをリアルタイムに取得する。ツイートテキストから放置自転車の台数を正規表現で抽出する。また、ツイートの緯度経度の情報を基に、放置自転車 LOD 内に存在する近くの観測地点を図 4.5 の SPARQL クエリを用いて放置自転車 LOD の SPARQL エンドポイント*7に問い合わせることで検索する。放置自転車 LOD はオープンソースの RDF ストアである Virtuoso*8に保存されており、SPARQL クエリ

*4 <https://developers.google.com/places/?hl=ja>

*5 <https://developer.foursquare.com/>

*6 <http://www.data.jma.go.jp/obd/stats/etrn/index.php>

*7 <http://www.ohsuga.is.uec.ac.jp/sparql>

*8 <http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/>

内で Virtuoso 固有の関数を使用することができる。図 4.5 では接頭辞「bif:」がこれに当たり、指定した緯度経度および半径で範囲検索することが可能である。本研究では半径を 25 メートルに設定している。半径 25 メートル以内に放置自転車 LOD に登録されている観測地点が存在しない場合は、新規観測地点として新たに追加するため、周辺の POI 情報を Google Places API と Foursquare API を使用して取得する。より多くの POI を取得するため 2 つの異なる API を使用している。取得した POI の内、最寄りの POI 名に「前」「付近」などの修飾語を付与したものを観測地点の名称とし、これを基に以下のように観測地点のリソースの URI を設計する。

`http://www.ohsuga.is.uec.ac.jp/bicycle/resource/obs_{観測地点名称}`

最寄り以外の POI は、観測地点の周辺施設として観測地点のリソースから GeoNames Ontology^{*9}で定義されている nearby プロパティでリンク付けする。GeoNames Ontology は地理空間情報のオントロジーである。周辺施設のリソースの URI は以下のように設計する。

`http://www.ohsuga.is.uec.ac.jp/bicycle/resource/{周辺施設名}`

しかし、Google Places API と Foursquare API では、同一の POI を指していても名称が異なる例がある。例えば、Google Places API で取得できる「ローソン 調布小島町二丁目」と Foursquare API で取得できる「ローソン 調布小島町二丁目店」は同じ店を指しているが名称が異なっている。そこで、編集距離を基にした式 4.9 により POI 名の類似度を計算し、類似度が閾値を超えた場合は同一 POI として統合する。本研究では閾値を 0.8 に設定している。 $EditDist(x, y)$ は文字列 x, y の編集距離であり、文字列 x を文字列 y に編集する際の削除・挿入コストを意味する。 $Max(x, y)$ は x, y の最大文字列長である。編集距離算出には Myers のアルゴリズム [33] を使用している。

$$sim(x, y) = 1 - \frac{EditDist(x, y)}{Max(x, y)} \quad (4.9)$$

Google Places API と Foursquare API では施設のカテゴリ情報を得ることが可能であり、本研究ではこのカテゴリと LinkedGeoData 内で定義されているクラスをマッピングしている。これにより、周辺施設リソースを LinkedGeoData 内のクラスのインスタンスとしている。しかし、タイプ情報が適切に付与されていない施設も存在するため、そのよ

^{*9} <http://www.geonames.org/ontology/documentation.html>

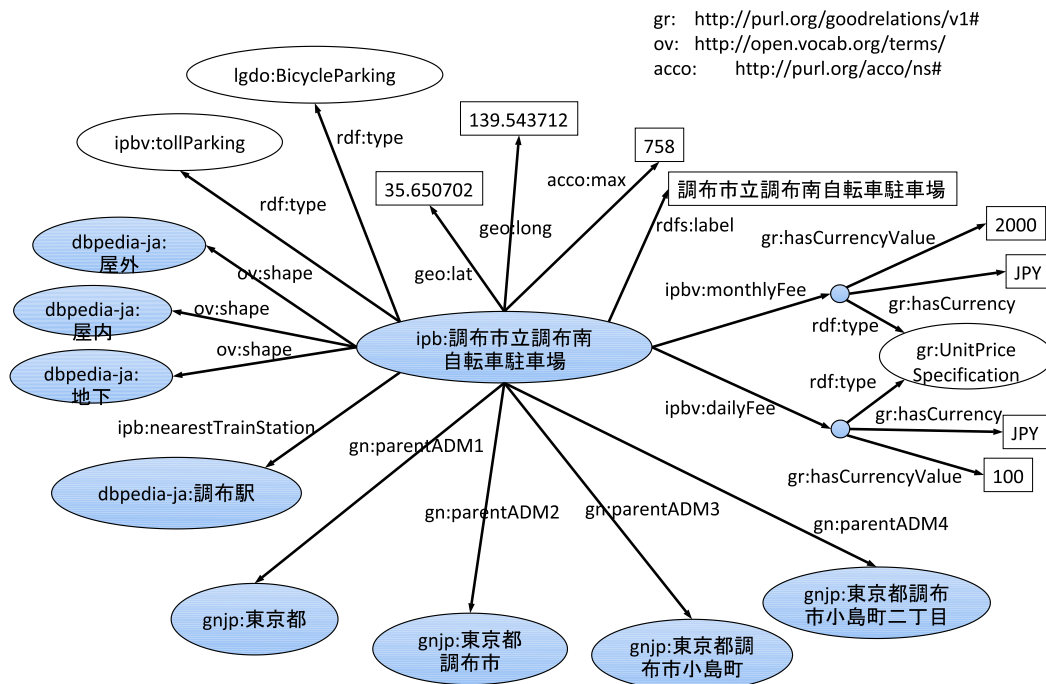


図 4.6 駐輪場の RDF グラフの例

うな施設に対しては施設名の文字列検索によりタイプを判定している。また、緯度経度情報を基に Yahoo!リバーズジオコード API^{*10}を使用して、最寄りの施設の住所を県名、町名などが分割された状態で取得する。この情報を基に、GeoNames.jp^{*11}へのリンクを生成する。GeoNames.jp は日本の地名の URI 基盤である。さらに、観測地点は最寄り駅として DBpedia Japanese のリソースとリンクし、最寄りの駐輪場として独自に設計したリソースとリンクしている。最寄り駅および最寄りの駐輪場に関しては既存のプロパティが見つからなかったため、独自に定義したプロパティを用いてリンクしている。また、現時点では駐輪場の LOD が存在していないため、本研究では市区町村の Web サイトから収集した情報を基に独自に駐輪場のリソースを構築している。観測地点は最寄りの駐輪場としてこのリソースとリンクする。図 4.6 に駐輪場の RDF グラフの例を示す。図 4.7 に観測地点と周辺施設の RDF グラフの例を示す。

放置自転車の観測時間は Time Ontology^{*12}の TemporalEntity クラスのインスタンス

*10 <http://developer.yahoo.co.jp/webapi/map/openlocalplatform/v1/reversegeocoder.html>

*11 <http://geonames.jp>

*12 <https://www.w3.org/TR/owl-time/>

wo: <https://www.auto.tuwien.ac.at/downloads/thinkhome/ontology/WeatherOntology.owl#>
 weather: <http://www.ohsuga.is.uec.ac.jp/weather/resource/>
 time: <http://www.w3.org/2006/time#>

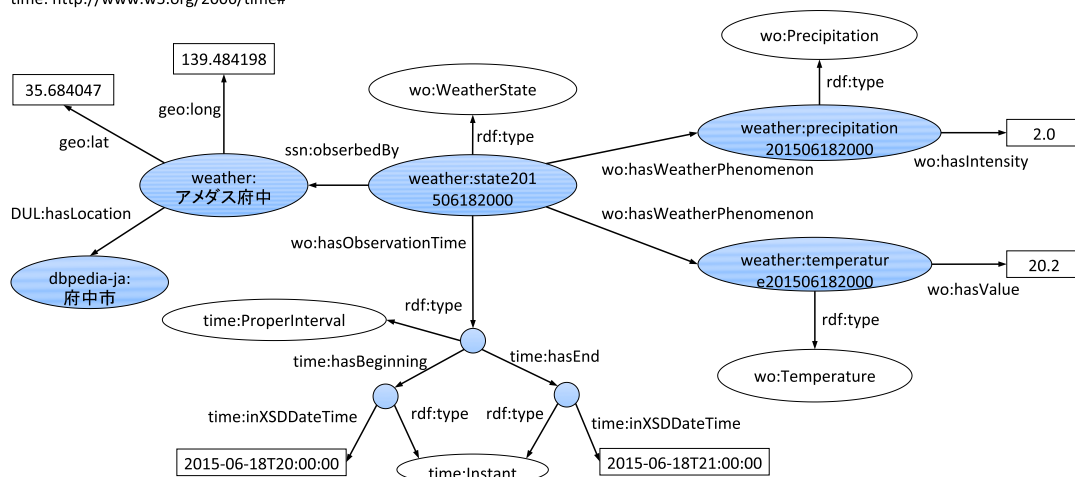


図 4.8 気象情報の RDF グラフの一部

気象情報のリソースは Weather Ontology^{*14} に基づいて、WeatherState クラスのインスタンスとして構築している。気象情報インスタンスは降水量インスタンスと気温インスタンスを持ち、これらのインスタンスが値を持っている。また、気象庁のアメダスによって観測された時間を Time Ontology に基づいた時間情報リソースとして持っている。放置自転車のリソースを対応する観測時間帯における、最寄りのアメダスで観測された気象情報のリソースに、独自に定義した ipb:weather プロパティでリンク付けする。図 4.8 に気象情報の RDF グラフの例を示す。

このように、ツイートをリアルタイムに取得した後に各種 API を用いて観測地点の情報を取得し、前述の LOD スキーマに従った RDF グラフを放置自転車 LOD に追加する。構築される放置自転車 LOD の一部を図 4.9 に示す。構築したデータ全体は RDF データベースに格納し、設置した SPARQL エンドポイント^{*15} を通して Web 上に公開している。

4.3.2 放置自転車 LOD の公開

データを LOD として Web 上に公開するためには、2.2 で述べた Linked Data の基本原則を満たしていなければならない。本研究で構築した放置自転車 LOD はこの基本原則を全て満たし、完全な LOD として公開している。放置自転車 LOD では、各リソースの URI

*14 <https://www.auto.tuwien.ac.at/downloads/thinkhome/ontology/WeatherOntology.owl>

*15 <http://www.ohsuga.is.uec.ac.jp/sparql>

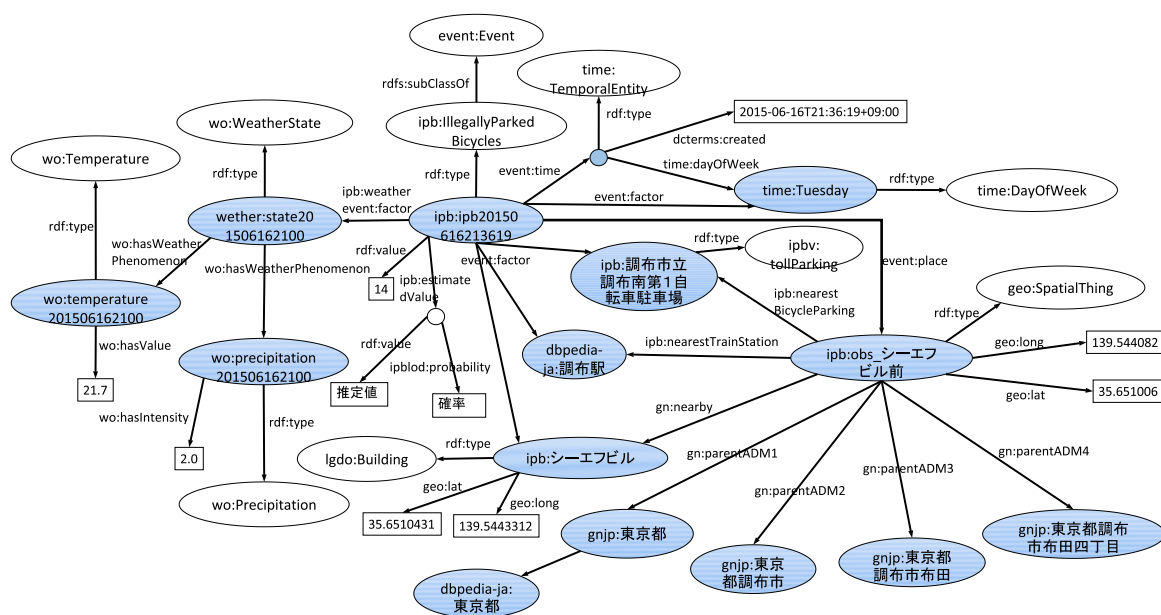


図 4.9 実際に構築される放置自転車 LOD の一部

にブラウザからアクセスすることで、人が見て理解できる情報、すなわち HR(Human Readable) ページを表示させる仕組みになっている。2015年8月17日21時42分26秒に調布駅前のアミューズランドモナコ前で観測された放置自転車の URI は以下のように定義される。

<http://www.ohsuga.is.uec.ac.jp/bicycle/resource/ipb20150817214226>

この URI にブラウザでアクセスすると、HTTP/1.1 の 303 リダイレクト (See Other) により以下の URI にリダイレクトされ、リダイレクト先の URI で図 4.10 のような HR ページが表示される。

<http://www.ohsuga.is.uec.ac.jp/bicycle/page/ipb20150817214226>

さらに、HTTP の Accept ヘッダでメディアタイプを指定することにより、様々な RDF シリアライゼーションでデータを取得することができる。上述の放置自転車リソースの URI に以下の様な curl^{*16} コマンドでアクセスすると、RDF/XML 形式でデータを取得することが可能である。

*16 <http://curl.haxx.se/>

```
curl -H "Accept: application/rdf+xml" http://www.ohsuga.is.uec.ac.jp/
bicycle/resource/ipb20150817214226
```

取得できる RDF/XML を以下に示す.

```
<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:event="http://purl.org/NET/c4dm/event.owl#"
  xmlns:ns3="http://www.ohsuga.is.uec.ac.jp/ipblod/vocabulary#" >
  <rdf:Description rdf:about="http://www.ohsuga.is.uec.ac.jp/bicycle/resource/ipb20150817214226">
  <rdf:type rdf:resource="http://www.ohsuga.is.uec.ac.jp/bicycle/resource/IllegallyParkedBicycles"
  />
  <rdf:value rdf:datatype="http://www.w3.org/2001/XMLSchema#int">23</rdf:value>
  <event:place rdf:resource="http://www.ohsuga.is.uec.ac.jp/bicycle/resource/obs_アミューズランドモナコ前" />
  <event:time rdf:nodeID="b621634" />
  <event:factor rdf:resource="http://www.w3.org/2006/time#Monday" />
  <event:factor rdf:resource="http://www.ohsuga.is.uec.ac.jp/bicycle/resource/EVE" />
  <event:factor rdf:resource="http://www.ohsuga.is.uec.ac.jp/weather/resource/府中市_state201508182100" />
  <event:factor rdf:nodeID="b621634"></event:factor>
  <ns3:weather rdf:resource="http://www.ohsuga.is.uec.ac.jp/weather/resource/府中市_state201508182100" />
  </rdf:Description>
</rdf:RDF>
```

このように、放置自転車 LOD は Linked Data の基本原則の 1 から 3 番目までを満たしている。さらに、DBpedia Japanese や LinkedGeoData などの外部 LOD ともリンクしているため、基本原則の 4 番目も満たしており完全な LOD として Web 上に公開している*17。

4.4 放置自転車台数推定による LOD 詳細化

放置自転車問題において、自転車の放置台数は実センサーではなく人によりソーシャルセンサ的に観測されるため、ある地点における全ての時間帯を観測することはできない。すなわち、構築した放置自転車 LOD には時間的な欠損が生じる。また、自転車が放置さ

*17 <http://www.ohsuga.is.uec.ac.jp/bicycle/dataset.html>

ipb20150817214226

Property	Value
event:factor	http://www.w3.org/2006/time#Monday
event:factor	http://www.ohsuga.is.uec.ac.jp/weather/resource/府中市_state201508182100
event:factor	http://www.ohsuga.is.uec.ac.jp/bicycle/resource/EVE
event:place	http://www.ohsuga.is.uec.ac.jp/bicycle/resource/obs_アミューズランドモナコ前
event:time	nodeID://b621634
ipb:weather	http://www.ohsuga.is.uec.ac.jp/weather/resource/府中市_state201508182100
rdf:type	http://www.ohsuga.is.uec.ac.jp/bicycle/resource/IllegallyParkedBicycles
event:factor	nodeID://b621634
rdf:value	23^^xsd:int

Raw Data in [CSV](#) | [RDF \(N-Triples N3 Turtle XML JSON\)](#) | [JSON-LD](#)

図 4.10 ブラウザで表示される HR ページの例

れる地点を全て把握しているわけではないため、実際には放置自転車が存在するがデータの取れていない地点、すなわち空間的な欠損も生じる。このような欠損データを推定・補完することにより、時空間的に密な LOD の構築が可能となり、データとしての有用性が高まる。そこで、本研究では、未観測日時の放置自転車の台数をベイジアンネットワークにより推定する。ベイジアンネットワークとはグラフィカルモデルの一種であり、因果関係を確率的に表現することで様々な推論を行う技術である。表 4.1 から、放置自転車は場所や時間などの属性により台数が変化することが分かっている。この因果関係をベイジアンネットワークにおける非循環有効グラフ構造として学習することで、確率的に台数を推定することが可能であると考えられる。しかし、これらの属性にも欠損が生じる可能性があり、そのままではベイジアンネットワークの学習が困難である。そこで、まず初めに欠損属性値を補完し、属性値に欠損のない学習データを入力として、ベイジアンネットワークにより未観測日時の放置自転車台数を推定する。推定結果をもとに RDF グラフを構築し、放置自転車 LOD に追加する。

4.4.1 欠損属性値の補完

まず、4.3 で構築した LOD から放置自転車の要因を取得する。放置自転車の要因は event:factor プロパティを辿ることで取得できるインスタンスまたは値と仮定し、これらをベイジアンネットワークにおける属性とする。使用する属性は、実際にデータとして取得することができた観測地点、曜日、時間帯、降水量、気温、駐輪場の料金(1日)、駐輪場の料金(1ヶ月)、周辺施設のタイプとした。ここで、周辺施設のタイプは表 4.1 の周辺施設に基づいて、Building, Bank, Games, DepartmentStore, Supermarket, Library, Police, School の 8 種類に限定した。なお、商店街、大規模店舗、公会堂等公的施設に関しては、Google Places API と Foursquare API でタイプを取得することができなかつたため、ここでは除いている。遊技場とパチンコ店はともに Games としている。しかし、観測地点によってはこれらの属性値にも欠損が生じている。そこで、構築した LOD から類似観測データを検索し、類似観測データの属性値を代入することで欠損属性値を補完する。観測データの類似度は Jaccard 係数により算出する。ここで観測データ o_i とは、全属性値に放置自転車の台数を加えた 15 次元のベクトル集合 O の要素を指す。各属性の集合は観測地点名 Loc 、曜日 Day 、時間帯 $Hour = \{1, \dots, 24\}$ 、降水の有無 $Rain = \{0, 1\}$ 、駐輪場の料金(1日) $DailyFee = \{0, 1, \dots\}$ 、駐輪場の料金(1ヶ月) $MonthlyFee = \{0, 1, \dots\}$ 、周辺施設のタイプ $Building, Bank, Games, DepartmentStore, Supermarket, Library, Police, School = \{0, 1\}$ とする。放置自転車の台数 Num は Jenks Natural Breaks[34] を用いて 4 つに分類し、 $Num = \{1, \dots, 4\}$ とする。台数の範囲は、0~7, 8~20, 21~42, 43~100 となっている。Jenks Natural Breaks は自然分類とも呼ばれ Geographic Information System (GIS) 分野でよく用いられている。すなわち、観測データ o_1, o_2 の類似度は式 4.10 により算出され、類似度の最も高い観測データから欠損属性値を補完する。

$$sim(o_1, o_2) = \frac{|o_1 \cap o_2|}{|o_1 \cup o_2|} \quad (4.10)$$

4.4.2 ベイジアンネットワークによる台数推定

4.4.1 の手法により得られた属性値に欠損の無いデータセットを入力データとして、4.4 初めで述べた時間的な欠損と空間的な欠損における放置自転車の台数をベイジアンネットワークにより推定する。本研究ではベイジアンネットワークのツールとして Weka^{*18}を使

*18 <http://www.cs.waikato.ac.nz/ml/weka/>

用した。入力データは要素数 15 のベクトル

$$o_i \in Loc \times Day \times Hour \times Rain \times DailyFee \times MonthlyFee \times Building \\ \times Bank \times Games \times DepartmentStore \\ \times Supermarket \times Library \times Police \times School \times Num$$

の集合 O であり、観測データ数は 813 件である。すなわち、813 件のデータを訓練データとして、既知の観測地点の未観測日時における放置自転車の台数を推定する。ベイジアンネットワークにおけるパラメータとして、探索アルゴリズム、確率計算アルゴリズム、マルコフブランケット分類器の有無、最大親ノード数、評価指標の項目を変更しながら実験を行い、最も推定精度の高くなるパラメータを採用する。

4.4.3 推定精度評価を通じたパラメータ設定

813 件のデータを入力データとして、ベイジアンネットワークによる放置自転車台数の推定を行った。パラメータとして、探索アルゴリズムとして HillClimber を使用、確率計算アルゴリズムとして SimpleEstimator を使用、マルコフブランケット分類器を使用、最大親ノード数を 2、評価指標を BAYES とした時に最も精度が高くなった。10 分割交差検証を行った結果、推定精度は 65.2% となった。表 4.3 に Confusion Matrix を示す。この結果を受けて、推定精度を更に向上させるためには属性の見直しが必要であると考え、属性数の追加と削減を行った。

まず、4.1.2 の LOD スキーマ設計の際に除外した項目について着目した。LOD スキーマ設計の際に、個人データ取得の問題から event:agent クラスを除いていたが、放置者の情報は放置自転車問題では重要な要因であると考え、そこで、個人データに代わり集団の情報として人口統計データを属性として使用した。人口統計は政府統計のポータルサイトである e-stat^{*19} で提供されている API 機能^{*20} を用いて取得した。提供されている統計情報の内、各市区町村における「可住地面積 1 km^2 当たり人口密度」と、「自宅外通勤・通学者数（鉄道・電車）」の項目のデータを利用した。この 2 つの属性を追加して再度 10 分割交差検証を行ったが、推定精度に変化は見られなかった。

次に、周辺施設のタイプに着目した。4.3 で述べた LOD 構築の際には、周辺施設のタイプに制限を設けていなかったが、ベイジアンネットワークによる台数推定の際には表 4.1 で抽出した周辺施設に限定した。この選択は、放置自転車の要件抽出の際に得られた

*19 <http://www.e-stat.go.jp/>

*20 <http://www.e-stat.go.jp/api/>

周辺施設のタイプのみが、放置自転車の要因に影響を与えるという仮説の元に行った。しかし、実際には放置自転車の要因として、これらの周辺施設意外も要因になりうるという新たな仮説を立て、周辺施設のタイプを制限せずに属性として追加した。その際、放置自転車 LOD 内に一度も現れない周辺施設のタイプは削除した。結果として周辺施設のタイプは 68 個に増加した。使用した周辺施設のタイプを以下に示す。

Pharmacy, Games, Park, Retail, Restaurant, AccountantOffice, University, Discount, ArtGallery, FastFoot, Library, BusStation, Prison, Platform, CarDealer, Gym, Parking, RealEstate, Church, Cafe, Supermarket, Hospital, NightClub, EmploymentAgencyOffice, Mall, Deli, Sport, Hairdresser, Doctor, Bar, Bank, Hackerspace, BookShop, Amenity, Clothing, Dentist, Hall, Electronics, Office, Motorbike, Food, Garden, BeautySalon, PlaceOfWorship, School, GovernmentOffice, Building, Spa, Health, InsuranceOffice, WineryShop, DepartmentStore, Cosmetics, Apparel, CarRental, TrainStation, VideoRental, Hotel, Cinema, Financial, MotorwayJunction, Medical, CoffeeShop, Convenience, Construction, TravelAgentOffice, Lawyer, Landmark

しかし、これらの周辺施設のタイプを属性として追加したところ、最大推定精度が 64.2% に下がる結果となった。これは属性の過多が原因と見られるため、周辺施設のタイプのクラスタリングすることで対応する。放置自転車 LOD では、周辺施設のタイプを LinkeGeoData のクラスとしているため、LinkedGeoData オントロジーにおける上位クラスを取得することができる。そこで、周辺施設のタイプに上位クラスが存在するリソースは、上位クラスを属性とすることでクラスタリングした。なお、Amenity クラスは非常に多くのクラスの上位クラスとなっているため、上位クラスが Amenity しか無いものは現タイプを使用するものとする。最終的に使用した周辺施設のタイプは以下に示す 35 個となった。

Pharmacy, Retail, Restaurant, Police, University, FastFood, BusStation, Gym, Parking, Church, Cafe, Supermarket, Hospital, Nightclub, Sport, Hairdresser, Doctor, Bar, Bank, Amenity, Dentist, Hall, Office, PlaceOfWorship, School, CommunityCentre, Building, Spa, CarRental, InternetCafe, CoffeShop, Construction, Layer, Leisure, Shop

この結果、探索アルゴリズムとして LAGDHillClimber を使用、確率計算アルゴリズムとして SimpleEstimator を使用、マルコフブランケット分類器を使用、最大親ノード数を 7、評価指標を BAYES とした時に、10 分割交差検証の精度が 66.7% と向上した。さらに、データの偏りを軽減するため全体から 9 割をランダムサンプリングし最大親ノード数を 3 に変更後、10 分割交差検証を計 10 回行った時の平均推定精度が 70.3% となり、本実験において最も高い結果を得た。詳細な結果は 5 章で記す。

表 4.3 属性およびパラメータ修正前の Confusion Matrix

	分類 1	分類 2	分類 3	分類 4
分類 1	345	59	5	2
分類 2	96	132	30	2
分類 3	12	63	52	2
分類 4	2	1	9	1
TP	0.839	0.508	0.403	0.077
加重平均	0.652			

```

PREFIX event: <http://purl.org/NET/c4dm/event.owl#>
PREFIX dcterms: <http://purl.org/dc/terms/>
RREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
REFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>
SELECT DISTINCT ?label ?date ?lat ?lng
FROM <http://www.ohsuga.is.uec.ac.jp/ipblod>
WHERE {
    ?s event:time ?time .
    ?time dcterms:created ?date .
    ?s event:place ?place .
    ?place rdfs:label ?label ;
        geo:lat ?lat;
        geo:long ?lng .
} order by ASC(?date)

```

図 4.11 観測地点と観測日時を全て取得する SPARQL クエリ

4.4.4 放置台数推定結果を用いた LOD の時系列補完

ベイジアンネットワークによる放置台数推定後に、推定結果とその確率を用いて RDF グラフを生成し、放置自転車 LOD に追加する。まず、放置自転車 LOD の SPARQL エンドポイントに図 4.11 に示す SPARQL クエリを送信し、全観測データを取得する。各観測地点において、最初に観測された日から最後に観測された日までの観測状況を全て調べ、午前 9 時または午後 9 時に観測がされていない場合に放置自転車の台数を推定する。ベイジアンネットワークの実行には Weka で提供されている Java API を使用し、4.4.3 のパラ

```

ipb:ipb_ガイア調布駅前ビル前_20150911090000 a ipb:IllegallyParkedBicycles ;
    event:factor <http://www.ohsuga.is.uec.ac.jp/weather/resource/府中市_state20150
9120900> , time:Friday ;
    event:place ipb:obs_ガイア調布駅前ビル前 ;
    event:time [ a time:DateTimeDescription ;
        dcterms:created "2015-09-11T09:00:00+09:00"^^xsd:dateTime ;
        time:day "11"^^xsd:decimal ;
        time:dayOfWeek time:Friday ;
        time:hour "9"^^xsd:decimal ;
        time:minute "0"^^xsd:decimal ;
        time:month "9"^^xsd:decimal ;
        time:second "0"^^xsd:decimal ;
        time:year "2015"^^xsd:decimal
    ];
    event:factor [ a time:DateTimeDescription ;
        dcterms:created "2015-09-11T09:00:00+09:00"^^xsd:dateTime ;
        time:day "11"^^xsd:decimal ;
        time:dayOfWeek time:Friday ;
        time:hour "9"^^xsd:decimal ;
        time:minute "0"^^xsd:decimal ;
        time:month "9"^^xsd:decimal ;
        time:second "0"^^xsd:decimal ;
        time:year "2015"^^xsd:decimal
    ];
    ipbv:estimatedValue [ rdf:value "0-7" ;
        ipbv:probability "0.7715675384230002"^^xsd:double
    ];
    ipbv:weather <http://www.ohsuga.is.uec.ac.jp/weather/resource/府中市_state20150
9120900> .

```

図 4.12 推定結果を基に構築される RDF (Turtle 形式)

メータ設定で未観測日時の放置自転車の台数を推定する。推定値とその確率を基に RDF を生成し、放置自転車 LOD に追加する。生成する RDF を Turtle 形式で図 4.12 に示す。

4.5 放置自転車 LOD の可視化

最後に、データを可視化することで、人間がデータ内容を直感的に理解できるようにする。特に、放置自転車のような都市問題に関するデータを可視化することにより、地域住民に問題意識を喚起することができ、さらなるデータの収集が見込まれる。

構築した放置自転車 LOD は Web 上に公開されており、SPARQL を受け付けるエンドポイントを設置しているため、誰でも利活用が可能な状態である。本研究では放置自転車 LOD 活用の一例として、市民の問題意識向上に向けて放置自転車を可視化する Web ア

アプリケーション「放置自転車マップ*²¹」を公開した。アプリケーションの動作例を 4.13 に示す。このアプリケーションでは、自転車の放置状況を地図上にアニメーション表示することができる。開始日時と終了日時を選択し、可視化の方法としてヒートマップまたはマークを選択し、再生ボタンを押すと放置自転車の時系列変化がアニメーション表示される。図 4.13 は Web アプリケーションに Android のスマートフォンでアクセスし、調布駅付近の放置自転車をヒートマップとマークでアニメーション表示させた画面である。図 4.14 はデスクトップ PC で表示させた画面である。本アプリケーションはレスポンシブデザインとなっているため、PC、スマートフォン、タブレットなど様々なデバイスでの表示に対応している。

本アプリケーションは、図 4.3 のツイート Web アプリケーションと同一サイト上にあり、ツイート後にすぐに可視化された情報を見ることが可能になっている。このように、ツイート後に放置自転車情報を可視化させてユーザに見せることにより、ユーザに直接的なフィードバックを与えて新たなデータの投稿を促進させる効果も期待できる。

本研究では、ソーシャルセンサから取得した放置自転車の情報だけでなく、ベイジアンネットワークによる欠損推定値も確率とともに LOD に含めており、時系列的に密な LOD となっている。そのため、本アプリケーションのように放置自転車の時系列変化を可視化する場合、欠損値補完前より視覚的に滑らかで動的な可視化が可能になる。図 4.15 に欠損値補完前と欠損値補完後の可視化アプリケーションの動作を示す。図 4.15 は調布駅前の同時刻におけるアニメーションの途中経過である。欠損値補完前の描写は時系列変化が断続的なのに対し、欠損値補完後のデータは時系列変化が連続的となっている。

*²¹ <http://www.ohsuga.is.uec.ac.jp/bicycle/>



図 4.13 可視化アプリケーションのスマートフォン画面例

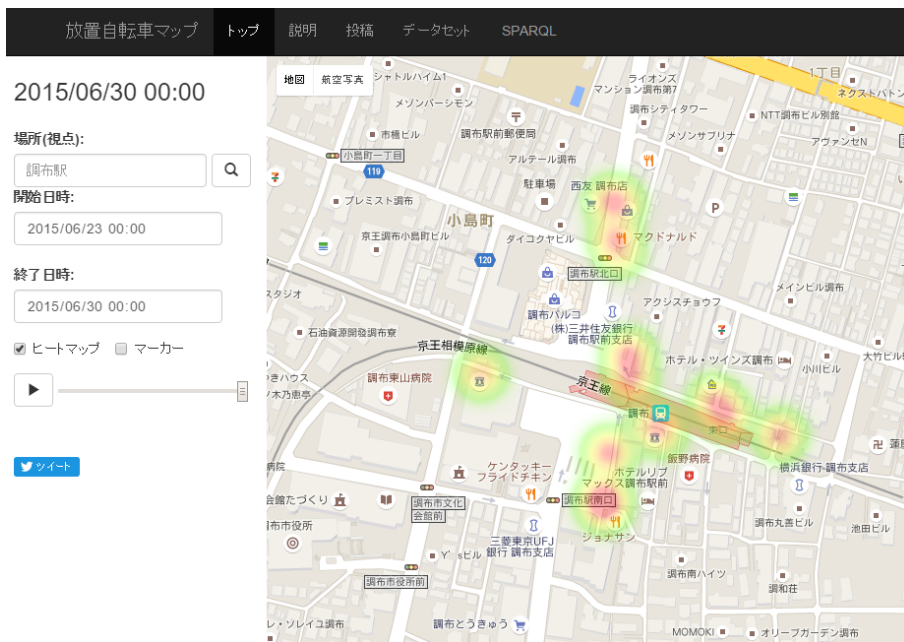
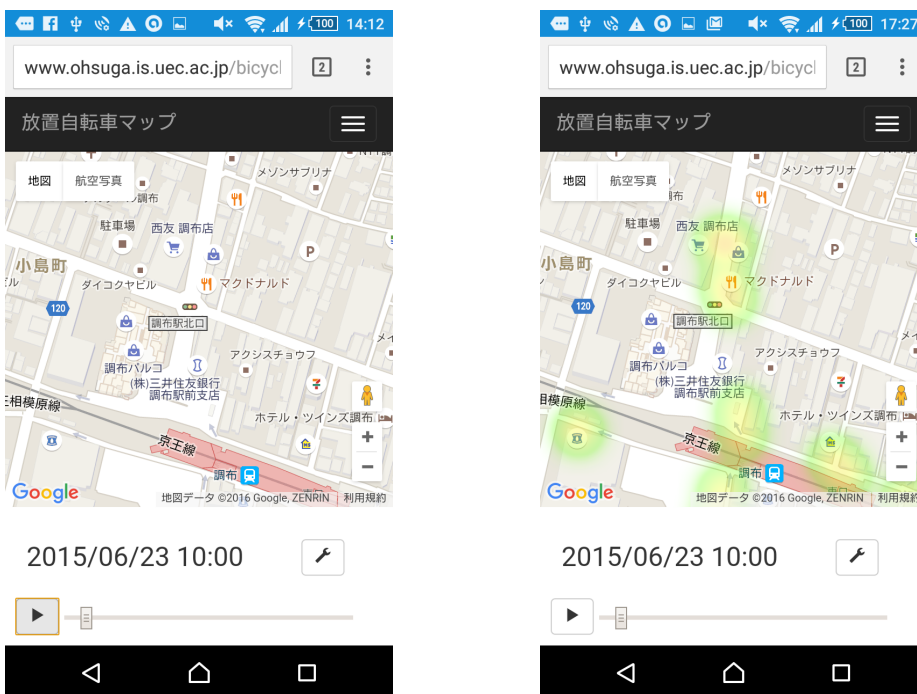


図 4.14 欠損値推定以前の可視化アプリケーションの動作例



欠測値(台数)推定前

欠測値(台数)推定後

図 4.15 欠損値補完前後の可視化アプリケーションの表示例

第 5 章

実験結果と考察

本章ではデータの収集と LOD 構築，ベイジアンネットワークによる放置台数推定の 2 つの観点から，それぞれの実験結果と考察を述べる。

5.1 データ収集と LOD 構築の結果

本研究で収集した一次データのうち，観測地点を特定できる使用可能な観測データは 813 件であった。表 5.1 に観測データの統計情報を示す。また，推定結果を追加した放置自転車 LOD のトリプル数は 219,804 となった。さらに，放置自転車 LOD は W3C の LinkingOpenData コミュニティプロジェクト^{*1}の作成する Linking Open Data cloud diagram(図 3.1) の加入条件を満たしている。Linking Open Data cloud diagram の加入条件は以下の 5 項目をすべて満たすことである。

1. 参照解決可能な URI である
2. コンテンツネゴシエーションの有りまたは無しで，RDF データを代表的なフォーマット (RDFa, RDF/XML, Turtle, N-triples 等) で参照できる
3. データセットは 1000 トリプル以上である
4. データセットは図内にあるデータセットへ 50 以上の RDF リンクを持っている
5. RDF クローリングか RDF ダンプか SPARQL エンドポイントによりデータセット全体にアクセスできる

放置自転車 LOD は図内に含まれている LOD として LinkedGeoData や GeoNames や DBpedia Japanese への RDF リンクを 50 以上含んでおり，その他の条件も全て満たして

^{*1} <https://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

いる。この結果、放置自転車 LOD はこれらの LOD を介して間接的に多くの LOD とリンクすることとなり、LOD としての有用性が向上した。今後放置自転車意外にも地域課題、社会課題の LOD が構築されることで、それらの LOD とリンクして新たな価値を生み出す可能性がある。

表 5.1 観測データの統計

地域	観測地点数	観測データ数
東京都調布市布田	17	634
神奈川県横浜市中区港町	1	37
東京都練馬区桜台	1	36
東京都練馬区豊玉上	1	31
東京都府中市府中町	7	18
東京都調布市小島町	2	17
東京都武蔵野市吉祥寺本町	2	8
東京都武蔵野市吉祥寺東町	2	4
東京都武蔵野市吉祥寺南町	1	4
北海道札幌市中央区北三条西	1	4
北海道札幌市中央区北四条西	2	3
神奈川県横浜市磯子区森	2	3
北海道札幌市中央区北一条西	2	3
東京都国分寺市本町	2	3
北海道札幌市北区北六条西	2	2
北海道札幌市中央区大通西	2	2
北海道札幌市中央区北五条西	1	1
東京都練馬区中村北	1	1
北海道札幌市中央区北二条西	1	1
東京都練馬区豊玉北	1	1

5.2 ベイジアンネットワークによる台数推定の結果

収集した観測データのうち属性値に欠損のある観測データは168件であり、まずこれを4.4.1の手法により補完した。欠損が生じていた属性は駐輪場料金の有無であった。次に属性値の欠損を全て補完した観測データを入力データとして、ベイジアンネットワークにより放置台数を推定した。

10分割交差検証を10回行った平均推定精度は70.3%であった。適合率、再現率、F値の平均値を表5.2に示す。実際に、このデータセットを訓練データとして、調布市の10箇所の観測地点の未観測日時の放置自転車の台数を予測し、その後現地で観測した正解データと比較した結果、7地点において台数が推定範囲内であった。Confusion Matrixを表5.1から表5.11に示す。結果として、観測データ数の少ない分類3(21~42台)や分類4(43~100台)の推定精度が低いため、全体としての精度を落とすことになったが、割合の大きい分類1(0~7台)の推定精度は高くなった。

ベイジアンネットワークの精度が上がらなかった原因として、観測データ数が十分でないことと不完全データであることが考えられる。今回使用したデータは全ての観測地点において均等に観測したデータではなく、観測地点や台数によってデータの量に違いがある。そのため、各条件付き確率を推定するのに十分なデータ数ではなかったと考えられる。特に台数の多いデータが少なく、正しく推定できていないという結果になった。今後は広範囲にデータ投稿協力を呼びかけ、特に放置自転車の多い地域のデータ収集が必要になると考える。

また、観測地点の範囲決定に個人差があったことも原因として考えられる。今回、放置自転車が一続きに集まっている地点を観測地点としたが、その範囲を明確に定めておらず、人によって台数の数え方に違いが生じていた。例えば自転車が数メートル置きに断続的に放置されている場合、それらを一つの観測地点における放置自転車とするか、あるいは別々の観測地点とするかという問題がある。現状では4.3.1で述べたように、半径25m以内の地点は同一観測地点として統合される。そのため、別々の観測地点として分けて報

表 5.2 10回の10分割交差検証における適合率、再現率、F値の平均値

適合率	再現率	F 値
0.694	0.703	0.698

告した場合に、それぞれの報告が同一観測地点における同一時間帯での台数報告となっていた。この結果、ベイジアンネットワークで台数推定を行う際に、同一観測地点における同一時間帯での少数台数の入力データが増加し、精度に影響を与えたと考えられる。今後は放置自転車集合の定義を明確化し、データ収集の際にも人によって台数カウント方法に差がでない手法を考案する必要がある。今後の課題として、図 4.3 のツイート Web アプリに、現在地点を中心とした観測対象範囲の円を表示する機能や、投稿された画像を解析して台数を推定する機能の実装考えている。

また、観測地点の周辺施設のタイプに関しては、Google Places API や Foursquare API により取得できなかったものが多く存在することも、精度を下げた原因として考えられる。本実験では API で取得可能なタイプと LinkedGeoData のクラスをマッピングしており、これに基づいて LOD を構築している。しかしながら、API 内で “establishment” や “point of interest” などの抽象的なタイプしか付与されていない施設が多く、ベイジアンネットワークの属性として追加するに至らなかった。このことも、精度を下げた原因として考えられる。本研究では施設名称に対して単純な文字列マッチングを行うことで、少量のタイプ未判別の施設に対してタイプの判別をしているが、さらなる改善の余地が見られる。

また、Web API や行政の Web サイトからだけでは取得することができなかった周辺施設情報も存在する。例えばスーパーや百貨店には専用の駐輪場を設置しているところがある。多くの場合、1 時間から数時間単位で課金される駐輪場であり、このような駐輪場の有無は近辺の放置自転車の台数に影響を与えたと考えられる。しかしながら、これらの民間駐輪場・専用駐輪場はデータとして取得することが困難であり、放置自転車 LOD 内に含まれていない専用駐輪場情報も多く存在する。今後はこれらの民間駐輪場・専用駐輪場の情報もクラウドソーシングで取得することを検討する。

さらに、今回の調査だけでは放置自転車の要因の全てを取得することはできていなかった可能性がある。例えば集客イベントや道路工事などは放置自転車の台数を左右する要因になりうると考えられる。しかし、4.1 で提案した LOD スキーマ設計の方法論に従い、検索エンジンから放置自転車の記事を検索した結果、これらのキーワードを発見することは無く、放置自転車の要因と断定することはできなかった。このように、対象の地域課題発生の要因として人が連想できるものを、LOD スキーマ設計の際に追加する方法論を組み込み、評価実験を行うことが必要であると考えられる。これについては今後の課題とさせていただきたい。

表 5.3 属性およびパラメータ修正後の Confusion Matrix (1)

	分類 1	分類 2	分類 3	分類 4
分類 1	321	44	1	3
分類 2	65	149	18	2
分類 3	15	46	53	2
分類 4	2	6	2	1
TP	0.870	0.637	0.457	0.091
加重平均	0.718			

表 5.5 属性およびパラメータ修正後の Confusion Matrix (3)

	分類 1	分類 2	分類 3	分類 4
分類 1	313	49	5	2
分類 2	64	151	19	0
分類 3	20	41	55	0
分類 4	2	7	2	0
TP	0.848	0.645	0.474	0.000
加重平均	0.711			

表 5.7 属性およびパラメータ修正後の Confusion Matrix (5)

	分類 1	分類 2	分類 3	分類 4
分類 1	311	53	4	1
分類 2	67	149	15	3
分類 3	20	56	38	2
分類 4	2	0	7	2
TP	0.843	0.637	0.328	0.182
加重平均	0.685			

表 5.9 属性およびパラメータ修正後の Confusion Matrix (7)

	分類 1	分類 2	分類 3	分類 4
分類 1	304	62	1	2
分類 2	67	153	13	1
分類 3	17	52	45	2
分類 4	3	3	2	3
TP	0.824	0.654	0.388	0.273
加重平均	0.692			

表 5.11 属性およびパラメータ修正後の Confusion Matrix (9)

	分類 1	分類 2	分類 3	分類 4
分類 1	320	41	8	0
分類 2	67	148	19	0
分類 3	15	42	59	0
分類 4	6	4	1	0
TP	0.867	0.632	0.509	0.000
加重平均	0.722			

表 5.4 属性およびパラメータ修正後の Confusion Matrix (2)

	分類 1	分類 2	分類 3	分類 4
分類 1	321	46	1	1
分類 2	72	144	17	1
分類 3	15	62	38	1
分類 4	1	3	7	0
TP	0.870	0.615	0.328	0.000
加重平均	0.689			

表 5.6 属性およびパラメータ修正後の Confusion Matrix (4)

	分類 1	分類 2	分類 3	分類 4
分類 1	319	43	6	1
分類 2	49	174	11	0
分類 3	18	39	58	1
分類 4	5	3	3	0
TP	0.864	0.744	0.500	0.000
加重平均	0.755			

表 5.8 属性およびパラメータ修正後の Confusion Matrix (6)

	分類 1	分類 2	分類 3	分類 4
分類 1	313	50	6	0
分類 2	86	133	14	1
分類 3	24	52	40	0
分類 4	4	5	0	2
TP	0.848	0.568	0.345	0.182
加重平均	0.668			

表 5.10 属性およびパラメータ修正後の Confusion Matrix (8)

	分類 1	分類 2	分類 3	分類 4
分類 1	322	42	5	0
分類 2	74	144	16	0
分類 3	14	51	51	0
分類 4	5	6	0	0
TP	0.873	0.615	0.440	0.000
加重平均	0.708			

表 5.12 属性およびパラメータ修正後の Confusion Matrix (10)

	分類 1	分類 2	分類 3	分類 4
分類 1	318	49	2	0
分類 2	85	136	13	0
分類 3	32	39	43	2
分類 4	5	4	1	1
TP	0.862	0.581	0.371	0.091
加重平均	0.682			

第6章

まとめと今後の課題

本論文では、放置自転車問題を解決を目的として、時間的・空間的に詳細な LOD を持続的に構築するシステムについて述べた。そのために、放置自転車 LOD に必要な要件を抽出した上でスキーマを設計し、これに基づいて、SNS や行政の Web サイトから収集した一次データを統合し LOD 化した。さらに、LOD スキーマ定義に基づいて属性選択したベイジアンネットワークを用いることで、未観測時間における放置自転車の台数を推定し、LOD に追加することで補完した。これにより、放置自転車に関して地理的に詳細かつ時間的に密な時系列 LOD を構築し、データの有用性を向上させた。データ活用の一例として放置自転車の分布を可視化するアプリケーションを公開し、視覚的に滑らかな表示を実現した。このような可視化により地域住民の問題意識が向上し、更なるデータの収集につながることを期待できる。

今後は5章で述べた課題を解決するため、可視化アプリケーションによるインセンティブを設計した上で、より広範囲にデータ収集協力を呼びかける。また、放置場所、範囲、放置時間、原動機付自転車やバイクのカウントなどの、放置自転車の定義を明確化し、ユーザによる放置自転車の台数カウント方法の統一化を図る。作業の省力化のため、写真から台数を推定することも検討している。周辺施設のタイプが判別できない課題に関しては、より多くの辞書を用いた文字列マッチングや類似度計算、あるいは機械学習を使用することを検討する。また、本研究で取得することができなかった放置自転車の要因を調査し、新たに LOD スキーマに追加することを検討する。さらに、この作業手順を LOD スキーマ設計の方法論として組み込むことも検討する。提案システムの導入による放置自転車 LOD の成長率およびインセンティブ付与による効果に関しては、長期的な評価が必要であるため今後の課題としたい。

本論文で述べた循環型 LOD 構築システムは、今後他の地域課題・社会課題にも適用で

きる可能性がある。これにより、様々な社会課題に関する詳細な LOD が構築されることで、単一の社会課題解決のみならず、社会課題を複合的に捉えてマッシュアップする社会課題解決型イノベーション^{*1}の創出が期待できる。

^{*1} <http://www2.deloitte.com/jp/ja/pages/strategy/articles/cbs/new-business-development1.html>

謝辞

本研究を行うにあたり、ご多忙の中、終始適切かつ丁寧なご指導をして下さった川村隆浩客員准教授、大須賀昭彦教授、田原康之准教授に深く感謝致します。清雄一助教にはご多忙の中、週1回のゼミを始めとして熱心な研究指導を賜り、貴重な勉学の機会を与えてくださったことに深く御礼申し上げます。また、研究の機会と議論・研鑽の場を提供して頂き、ご指導頂いた国立情報学研究所/東京大学の本位田真一教授をはじめ活発な議論と貴重なご意見を頂いた研究グループの皆様、大須賀・田原研究室の皆様にご感謝の意を表します。さらに、本研究を行う上で必要なデータ収集に協力して下さった Twitter ユーザの皆様、データを公開している政府組織と地方公共団体、Linked Open Data の普及活動に協力して下さった Linked Open Data チャレンジ実行委員会の皆様、特定非営利活動法人 リンクト・オープン・データ・イニシアティブの皆様にご感謝の意を表します。最後に、いつも暖かく応援して下さった家族、親戚の皆様にご心より感謝申し上げます。

参考文献

- [1] S. Bischof, C. Martin, A. Polleres, and P.Schneider. Collecting, integrating, enriching and republishing open city data as linked data. *Proceedings of the 14th International Semantic Web Conference (ISWC)*, pp. 57–75, 2015.
- [2] 溝口理一郎, 人工知能学会. オントロジー工学. 2005.
- [3] Aldo Gangemi, Nicola Guarino, Claudio Masolo, Alessandro Oltramari, and Luc Schneider. Sweetening ontologies with dolce. *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management*, pp. 166–181, 2002.
- [4] Riichiro Mizoguchi. Yamato: yet another more advanced top-level ontology. *Proceedings of the Sixth Australasian Ontology Workshop*, pp. 1–16, 2010.
- [5] Michael Ashburner and et.al. Gene ontology: tool for the unification of biology. *Nature Genetics*, Vol. 25, No. 1, pp. 25–29, 2000.
- [6] 大江和彦. 病名用語の標準化と臨床医学オントロジーの開発. *情報管理*, Vol. 52, No. 12, pp. 701–709, 2010.
- [7] 來村徳信, 溝口理一郎. 故障オントロジー. *人工知能学会誌*, Vol. 14, No. 5, pp. 828–837, 1999.
- [8] Claus Stadler, Jens Lehmann, Konrad Höffner, and Sören Auer. Linkedgeodata: A core for a web of spatialopen data. *Semantic Web Journal*, Vol. 3, No. 4, pp. 333–354, 2012.
- [9] A. Callahan, J. Cruz-Toledo, P. Ansell, and M.Dumontier. Bio2rdf release 2: Improved coverage, interoperability and provenance of life science linked data. *Proceedings of the 10th Extended Semantic Web Conference (ESWC)*, pp. 200–212, 2013.
- [10] Michel Dumontier, Alison Callahan, Jose Cruz-Toledo, Peter Ansell, Vincent Emonet, Francois Belleau, and Arnaud Droit. Bio2rdf release 3: A larger, more

- connected network of linked data for the life sciences. *Proceedings of the 13th International Semantic Web Conference (ISWC) Posters & Demonstrations Track*, pp. 401–404, 2014.
- [11] Fumihiko Kato, Hideaki Takeda, Seiji Koide, and Ikki Ohmukai. Building dbpedia japanese and linked data cloud in japanese. *Proceedings of the Joint International Workshop: 2013 Linked Data in Practice Workshop (LDPW2013) and the First Workshop on Practical Application of Ontology for Semantic Data Engineering (PAOS2013)*, pp. 1–11, 2013.
- [12] Fuyuko Matsumura, Iwao Kobayashi, Fumihiko Kato, Tetsuro Kamura, Ikki Ohmukai, and Hideaki Takeda. Producing and consuming linked open data on art with a local community. *Proceedings of the Third International Workshop on Consuming Linked Data (COLD2012)*, 2012.
- [13] 江上周作, 川村隆浩, 藤井章博, 大須賀昭彦. Bom エージェントの実現に向けた lod の構築. 電子情報通信学会論文誌, Vol. J98-D, No. 6, pp. 992–1004, 2015.
- [14] Luis von Ahn. Games with a purpose. *IEEE Computer*, Vol. 39, No. 6, pp. 92–94, 2006.
- [15] Matthew Zook, Mark Graham, Taylor Shelton, and Sean Gorman. Volunteered geographic information and crowdsourcing disaster relief: A case study of the haitian earthquake. *World Medical & Health Policy*, Vol. 2, No. 2, pp. 7–33, 2010.
- [16] 鈴木友基, 梶克彦, 河口信夫. クラウドソーシングによる屋内構造地図情報の構築と収集. 電子情報通信学会技術研究報告. MoMuC, モバイルマルチメディア通信, Vol. 111, No. 296, pp. 1–6, 2011.
- [17] 東田圭介, 櫻木伸幸. クラウドソーシングを用いた poi 情報収集. 電子情報通信学会技術研究報告. AI, 人工知能と知識処理, Vol. 111, No. 447, pp. 17–19, 2012.
- [18] Irene Celino, Simone Contessa, Marta Corubolo, Daniele Dell’ Aglio, Emanuele Della Valle, Stefano Fumeo, Thorsten Krüger. Linking smart cities datasets with human computation - the case of urbanmatch. *Proceedings of the 11th International Semantic Web Conference (ISWC)*, pp. 34–49, 2012.
- [19] Irene Celino. Geospatial dataset curation through a location-based game. *Semantic Web Journal*, Vol. 5, No. 7, 2014.
- [20] Edith Law and Luis von Ahn. Input-agreement: a new mechanism for collecting data using human computation games. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1197–1206, 2009.

- [21] Nicolas Kaufmann and Thimo Schulze. Worker motivation in crowdsourcing and human computation. *Proceedings of the AAAI Workshop on Human Computation (HCOMP)*, 2011.
- [22] Amit Sheth Harshal Patni, Cory Henson. Linked sensor data. *2010 International Symposium on Collaborative Technologies and Systems*, pp. 362–370, 2010.
- [23] Ghislain Ateazing, Oscar Corcho, Daniel Garijo, José Mora, María Poveda-Villalón, Pablo Rozas, Daniel Vila-Suero, and Boris Villazón-Terrazas. Transforming meteorological data into linked data. *Semantic Web Journal*, Vol. 4, No. 3, pp. 285–290, 2013.
- [24] Lihua Zhao, Ryutaro Ichise, Seiichi Mita, and Yutaka Sasaki. An ontology-based intelligent speed adaptation system for autonomous cars. *Proceedings of the 4th Joint International Semantic Technology conference (JIST)*, pp. 397–413, 2014.
- [25] Lihua Zhao, Ryutaro Ichise, Seiichi Mita, and Yutaka Sasaki. Ontologies for advanced driver assistance system. 人工知能学会セマンティックウェブとオントロジー研究会, Vol. SIG-SWO-035-03, , 2015.
- [26] 市瀬龍太郎. ストリーム推論. 人工知能学会誌, Vol. 30, No. 5, pp. 574–579, 2015.
- [27] Davide Francesco Barbieri, Daniele Braga, Stefano Ceri, Emanuele Della Valle, and Michael Grossniklaus. C-sparql: Sparql for continuous querying. *Proceedings of the 18th International Conference on World Wide Web (WWW)*, pp. 1061–1062, 2009.
- [28] Vanessa Lopez, Spyros Kotoulas, Marco Luca Sbodio, Martin Stephenson, Aris Gkoulalas-Divanis, and Pól Mac Aonghusa. C-sparql: Sparql for continuous querying. *Proceedings of the 11th International Semantic Web Conference (ISWC)*, pp. 148–163, 2012.
- [29] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives. Dbpedia: a nucleus for a web of open data. *Proceedings of the 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference (ISWC/ASWC)*, pp. 722–735, 2007.
- [30] R. Mizoguchi, M. Ikeda, K.Seta, and J. Vanwelkenhuysen. Ontology for modeling the world from problem solving perspectives. *Proceedings of the 1995 International Joint Conference on AI (IJCAI) Workshop on Basic Ontological Issues in Knowledge Sharing*, pp. 1–12, 1995.
- [31] 石川誠一, 久保成毅, 古崎晃司, 來村徳信, 溝口理一郎. タスク・ドメインロールに基づくオントロジー構築ガイドシステムの設計と開発石油精製プラントを例として. 人工

- 知能学会論文誌, Vol. 17, No. 5, pp. 585–597, 2002.
- [32] Soyoung Hwang and Donghui Yu. Gps localization improvement of smartphones using built in sensors. *International Journal of Smart Home*, Vol. 6, No. 3, pp. 1–8, 2012.
- [33] Eugene W. Myers. An $o(nd)$ difference algorithm and its variations. *Algorithmica*, Vol. 1, No. 1-4, pp. 251–266, 1986.
- [34] G.F. Jenks. The data model concept in statistical mapping. *International yearbook of cartography*, Vol. 7, No. 1, pp. 186–190, 1967.

研究業績

論文誌

1. **江上周作**, 川村隆浩, 清雄一, 田原康之, 大須賀昭彦: 社会課題解決に向けた循環型 LOD 構築システムの提案, 人工知能学会論文誌, 2016 年 (投稿中)
2. **Shusaku Egami**, Takahiro Kawamura, Yuichi Sei, Yasuyuki Tahara, and Akihiko Ohsuga: A Solution to Visualize Open Urban Data for Illegally Parked Bicycles, Transactions on Large-Scale Data and Knowledge-Centered Systems, 2016 (採録決定済み)
3. **江上周作**, 川村隆浩, 藤井章博, 大須賀昭彦: BOM エージェントの実現に向けた LOD の構築, 電子情報通信学会論文誌, Vol.J98-D, No.6, pp.992-1004, 2015 年 6 月

国際会議

1. **Shusaku Egami**, Takahiro Kawamura, Yuichi Sei, Yasuyuki Tahara, and Akihiko Ohsuga: Visualization of Open Urban Data for Illegally Parked Bicycles, CompleXity: Technology for Complex Urban Systems in The 49th Hawaii International Conference on System Sciences (HICSS-49), Jan 2016. (ロング発表)
2. **Shusaku Egami**, Takahiro Kawamura, Akihiro Fujii and Akihiko Ohsuga: Building of Industrial Parts LOD for EDI - A Case Study -, The 4th Joint International Semantic Technology Conference(JIST 2014), pp.146-161, Nov 2014. (ロング発表)

査読付き国内シンポジウム・ワークショップ

1. 江上周作, 川村隆浩, 清雄一, 田原康之, 大須賀昭彦: 社会課題解決に向けた循環型オープンデータ構築システムの提案, 合同エージェントワークショップ&シンポジウム 2015 (JAWS2015), pp.274-281, 2015年10月. (ショート発表)
2. 江上周作, 川村隆浩, 藤井章博, 大須賀昭彦: BOM エージェントの実現に向けた LOD の構築, 合同エージェントワークショップ&シンポジウム 2014 (JAWS2014), pp.115-118, 2014年10月. (ショート発表)

研究会

1. 江上周作, 川村隆浩, 清雄一, 田原康之, 大須賀昭彦: 放置自転車問題解決に向けた循環型 LOD 構築システムの提案, 第38回人工知能学会セマンティックウェブとオントロジー研究会, SIG-SWO-038-08, 2016年2月.

招待講演

1. 江上周作: 工業分野における Linked Open Data 活用に向けた取り組み, 参加型オープンデータで日本を元気にするシンポジウム - Linked Open Data チャレンジ 2015 キックオフ -, 2015年9月

受賞

1. 江上周作, 川村隆浩, 清雄一, 田原康之, 大須賀昭彦: 放置自転車 LOD, Linked Open Data チャレンジ Japan 2015 データセット部門 最優秀賞, 2016年2月