都市問題の分析に向けたナレッジグラフの構築と時間的・空間的・課題横断的な拡充手法の開発

江上 周作

電気通信大学大学院情報理工学研究科
博士（工学）の学位申請論文

2019 年 3 月

都市問題の分析に向けたナレッジグラフの構築と時間的・空間的・課題横断的な拡充手法の開発

博士論文審査委員会

主査　大須賀 昭彦　教授
委員　川村 隆浩　客員准教授
委員　田中 健次　教授
委員　西野 哲朗　教授
委員　柏原 昭博　教授

著作権所有者

江上 周作

2019

Constructing and Expanding Spatial, Temporal, and Cross-sectional Knowledge Graphs for Analyzing Urban Problems

Shusaku Egami

Abstract

Local governments have social problems closely related to urban spatial features and human behaviors, such as littering, graffiti, homeless, and illegally parked bicycles. In recent years, data-driven approaches are attracted for solving such urban problems, and each local government has begun to publish various data on the Web, such as statistical information and minute books. However, fine-grained data on urban problems is not only closed but also not accumulated; thus, it is insufficient to analyze data for solving those problems. For example, in terms of illegally parked bicycles, despite that there are several wide-area statistics per year, it is difficult to analyze for identifying the reason and proposing countermeasures based on such coarse data. Also, in terms of littering on streets, local governments cannot understand the current situation.

Moreover, since various factors are socially intertwined, urban problems are difficult to solve without understanding the causal relations among them. However, there is no data set including the urban problems' causality; thus, local governments cannot understand the current situation of urban problems occurring in the local areas.

In this study, we construct a spatio-temporal and cross-sectional data infrastructure for analyzing urban problems, then we aim to support for solving urban problems through indicating the analytic examples using the data infrastructure. In this study, we accumulated the urban problem data as knowledge graphs (KG), and published on the Web as Linked Open Data (LOD).

First, we first designed schemata of urban problems and collected data using social sensors, and then constructed the KGs based on the designed schemata. We then estimated

and complemented the spatio-temporal missing data. In terms of the temporal missing data, we estimated the data using bayesian networks based on the causal relations from the factors. In terms of the spatial missing data, we estimate the data using graph embedding methods and neural networks based on the results of the Computational Fluid Dynamics (CFD) simulation and geospatial features. In addition, we developed a web application that visualizes the constructed KGs, and introduced the application on the website of Office for Youth Affairs and Public Safety of the Tokyo Metropolitan Government through discussions with Tokyo Metropolitan Government.

Next, we constructed a cross-sectional KG including urban problems' causality and local governments' budgets. This is an upper-level KG of the fine-grained urban problem KGs. We first designed schema of the KG, then constructed the KG based on the words extracted from Web pages, SNS, open government data, and academic papers, using natural language processing methods and crowdsourcing. In addition, we showed several examples of searching causal relations of urban problems and the related local government's projects using the constructed KG. Moreover, we complemented causal missing links between entities that appeared in different contexts, using Semantic Web Rule Languages (SWRL) rules. Furthermore, we detected vicious cycles and root problems using SPARQL Protocol and RDF Query Language (SPARQL), then evaluated them based on comments from urban-problem experts in cooperation with the Osaka City Citizens Bureau.

We believe that this study helps to solve urban problems through the realization of the data infrastructure using the proposed construction and expansion methods of KGs. In the future, we will continue to expand the KGs over the long-term, and we aim to apply the KGs to the discussions in the local governments and NPOs.

都市問題の分析に向けたナレッジグラフの構築と
時間的・空間的・課題横断的な拡充手法の開発

江上 周作

概要

国内ではゴミのポイ捨て，放置自転車，ホームレス，犯罪など，都市の地理空間的な特性や人の行動に関係の深い様々な社会課題（都市問題）が発生している．近年，こうした都市問題のデータ利活用による課題解決が注目されており，政府を始め各自治体でも，統計情報や会議録など様々なデータのオープンデータ化の取り組みが進んでいる．しかし，都市問題そのものの詳細なデータに関しては，公開だけでなくデータの蓄積もそれ程進んでおらず，直接的なデータ分析を行うには十分でない．例えば放置自転車に関して言えば，自治体によっては年単位かつ広域な統計データは存在するものの，このような粒度の粗いデータをもとに，具体的な原因特定や解決施策の実施に向けた分析を行うことは困難である．また，ゴミのポイ捨て問題に至っては自治体がデータを収集することができておらず，一部自治体が企業に調査委託した結果を保持する程度であり，オープンデータとしては公開されていない．

また，都市問題は様々な原因が複雑に絡み合っている．そのため，解決施策実行に先立って自治体の抱える各種都市問題の原因や影響，都市問題同士の関係性を俯瞰して把握することや，解決に向けた費用対効果を把握することが望ましい．しかしながら，このような都市問題間の関係が整理されたデータセットは存在しておらず，自治体によっては担当地域において生じている都市問題を把握できていないこともある．

本研究では前述の背景を受けて，都市問題解決に向けた時間的・空間的・課題横断的なデータ分析活用基盤を構築し，活用例を示すことで問題解決の一助となることを目的とする． 本研究では都市問題に関するデータを知識グラフ（ナレッジグラフ）として蓄積し，これを Linked Open Data（LOD）の形式で公開した．

初めに，都市問題ごとにデータを収集して時間的・空間的に詳細なナレッジグラフを構築した．まず，ナレッジグラフを構築する上で重要となるデータ構造（スキーマ）を設計し，

次にソーシャルセンサを活用してデータを収集し，設計したスキーマに基づいてナレッジグラフを構築した．しかし, ソーシャルセンサにより集められるデータには時間的または空間的な欠損が生じる．そこで，本研究では都市問題に関するナレッジグラフの時間的な欠損値について，ベイジアンネットワークを用いて推定した．空間的な欠損値について，数値流体力学による都市の気流シミュレーション結果と地理空間情報を基に，Graph Embedding 手法とニューラルネットワークを用いて推定する手法を提案した．さらに，構築したナレッジグラフを活用して可視化アプリケーションを開発し，課題解決に向けて東京都との議論を通じて，青少年・治安対策本部のホームページで紹介した．

次に，複数の都市問題間の因果関係と関連予算情報に関するデータを収集し，課題横断的なナレッジグラフを構築した．これは，各都市問題の時空間的に詳細なナレッジグラフの上位概念的位置づけである．まず，スキーマを設計し，次に Web から収集した記事，SNS 情報，自治体オープンデータ，学術情報などを基に，自然言語処理とクラウドソーシングを用いて都市問題の因果関係を抽出する手法を提案した．設計したスキーマに基づいて，抽出した都市問題の因果関係や，関連する自治体事業予算情報を記述するナレッジグラフを構築した．構築したナレッジグラフを用いて，都市問題の因果関係と関連予算情報の検索例を示した．また，異なる文脈で出現したエンティティ同士の因果的な欠損について，セマンティック Web ルール言語により推論することで補完した．さらに，都市問題の因果関係からなる悪循環および悪循環を引き起こす根源的課題を SPARQL Protocol and RDF Query Language (SPARQL) で発見する実験を行い，大阪市市民局協力のもとで，都市問題の専門家による評価を行った．

本研究で示したナレッジグラフの構築と拡充手法により，データ分析活用基盤の実現を通して都市問題解決の一助となることが期待できる．今後，長期的なデータ拡充を続け，自治体や NPO での議論，市民協働イベント等での活用を通して社会実装していくことが課題である．

目 次

第 1 章　序論 **1**

1.1　本研究の背景 1

1.2　本研究の目的と意義 2

第 2 章　関連研究 **7**

2.1　ナレッジグラフと Linked Data の概要 7

2.1.1　ナレッジグラフ 7

2.1.2　Linked Data 8

2.1.3　オントロジー 9

2.2　社会課題および都市分析に関するナレッジグラフの研究 11

2.3　ナレッジグラフの構築に関する研究 12

2.4　ナレッジグラフの欠損補完と機械学習への適用 13

第 3 章　都市問題ごとのナレッジグラフの設計と構築 **17**

3.1　都市問題に関するナレッジグラフのスキーマの設計 17

3.1.1　ドメイン要件の抽出 18

3.1.2　クラス・インスタンス・プロパティの設計 20

3.2　都市問題の時空間的に詳細なデータの収集 21

3.3　ナレッジグラフの構築と公開 23

3.3.1　時系列データに基づくナレッジグラフのリアルタイム構築 23

3.3.2　非時系列データに基づくナレッジグラフの構築 25

第 4 章　都市問題ごとのナレッジグラフの時空間的な拡充と分析 **29**

4.1　ベイジアンネットワークを用いた時間的な欠損値の推定 29

4.1.1 欠損属性値の補完 30
4.1.2 ベイジアンネットワークによる放置自転車台数の推定 30
4.1.3 推定精度評価を通じたパラメータ設定 31
4.1.4 評価 33
4.1.5 考察 35
4.2 流体シミュレーションを用いた空間的な欠損値の推定 37
4.2.1 流体シミュレーションを用いたよどみ点の検出 38
4.2.2 セマンティクスを用いたよどみ点のフィルタリング 39
4.2.3 評価 42
4.2.4 考察 46
4.3 Graph Embedding とニューラルネットワークによる空間的欠損値の推定 . 47
4.3.1 流体シミュレーション結果と地理空間情報の Graph Embedding による推定 49
4.3.2 流体シミュレーション結果画像の CNN による推定 54
4.3.3 CNN と Graph Embedding を組み合わせた推定 55
4.3.4 評価 57
4.3.5 考察 63
4.4 都市問題に関するナレッジグラフの活用例の検討 67
4.4.1 SPARQL を用いたナレッジグラフの活用例 67
4.4.2 ナレッジグラフを活用した可視化アプリケーション 69

第 5 章 都市問題の課題横断的なナレッジグラフの構築と分析 77

5.1 都市問題の因果関係と自治体予算情報を含むナレッジグラフの設計と構築 . 78
5.1.1 都市問題の因果関係と自治体予算情報のスキーマ設計 78
5.1.2 データ収集と係り受け解析による因果関係の半自動的な抽出 80
5.1.3 クラウドソーシングを用いた因果単語候補のフィルタリング 81
5.1.4 因果単語抽出結果に基づくナレッジグラフの構築 82
5.1.5 自治体予算情報のナレッジグラフの構築 84
5.1.6 都市問題の因果関係抽出結果の考察 86

5.2 都市問題の課題横断的なナレッジグラフの拡充と分析 88
5.2.1 推論ルールを用いた欠損因果リンクの補完 88
5.2.2 都市問題解決施策の検討に向けたSPARQL 検索 93
5.2.3 都市問題の因果関係からなる悪循環発見の試み 95
5.2.4 悪循環につながる根源的課題の発見 96
5.2.5 評価 97
5.2.6 考察 98

第 6 章 結論 **101**
6.1 まとめ 101
6.2 今後の課題 103

謝辞 **107**

参考文献 **109**

研究業績 **119**

図 目 次

1.1 本研究の取り組み範囲 3
1.2 本研究の全体図 4

2.1 Linked Open Data cloud diagram (2017 年 8 月時点) 10

3.1 都市問題のデータ分析活用基盤のイメージ図 17
3.2 Event Ontology (http://purl.org/NET/c4dm/event.owl) 19
3.3 インスタンス加えて設計した LOD スキーマ 22
3.4 ツイート Web アプリの動作画面 24
3.5 一次データ収集と LOD 化の手順 25
3.6 実際に構築される放置自転車 LOD の一部 26
3.7 実際に構築されるポイ捨て調査結果の LOD の一部 28

4.1 調布駅周辺の 3D 地図 39
4.2 調布駅周辺の格子生成図 40
4.3 調布駅周辺の平均風速 41
4.4 本研究で定義したよどみ点のパターン 42
4.5 調布駅周辺にけるよどみ点の検出結果 43
4.6 調布駅周辺にけるフィルタリング後のよどみ点（最終的な推定値） 44
4.7 ベースライン手法 45
4.8 空間的欠損値の画像ベース推定手法とグラフベース推定手法の流れ 49
4.9 両国駅を中心とした格子生成図 50
4.10 両国駅の格子点における平均風速のプロット図 51
4.11 流体シミュレーション結果と地理空間情報のナレッジグラフ化のイメージ . 52

4.12 Graph Embedding 後にゴミの個数の予測に使用する全結合ニューラルネットワークのモデル 53
4.13 両国駅の CFD シミュレーション結果のヒートマップ表示画像 54
4.14 ゴミの個数の予測に使用する CNN のモデル 55
4.15 ゴミの個数の予測に使用する画像-グラフ結合手法のモデル 56
4.16 グラフベース手法の学習過程 58
4.17 画像ベース手法の学習過程 59
4.18 画像-グラフ結合手法の学習過程 60
4.19 セル内のゴミの個数の分布 61
4.20 セル内のゴミの個数の分布（0 を抜いた場合） 63
4.21 全観測地点における一時間ごとの平均放置自転車台数 68
4.22 可視化したアプリケーションの画面例 70
4.23 可視化したアプリケーションの画面例 71
4.24 空間的欠損値の可視化 72
4.25 可視化アプリケーションのページビュー数 73

5.1 2 種類のナレッジグラフをまたぐ分析のイメージ図 77
5.2 都市問題の因果関係と予算情報を表す Linked Data スキーマ 79
5.3 郊外型犯罪の要因のワードクラウド 82
5.4 最終的に構築される都市問題因果関係 LOD の一部 86
5.5 推論ルールにより補完される欠損因果リンク 90
5.6 ホームレスの因果関係の部分グラフ 93
5.7 ポイ捨ての因果関係の部分グラフ 94
5.8 根源的課題のグラフパターン 97
5.9 専門家による都市問題の因果関係評価の様子 99

表 目 次

3.1 抽出したキーワードのクラスタリング結果 20

4.1 観測データの統計 34
4.2 Confusion Matrix 35
4.3 各よどみ点パターンの精度 42
4.4 ベースライン手法と提案手法 1 の推定精度 46
4.5 ベースライン手法と提案手法 2 の推定精度 47
4.6 ポイ捨てゴミの個数の推定手法の精度比較 57
4.7 データの偏りを考慮した精度比較 62
4.8 放置自転車観測地点の近くの POI トップ 10 69
4.9 市区町村ごとのユーザ数トップ 10 74

5.1 因果単語抽出結果 88
5.2 Landis らによる Kappa 係数の解釈 88
5.3 自由記述欄の回答の内訳 89
5.4 発見された悪循環の数 96

第1章　序論

本章では，本研究の背景を述べた後，本論文の目的と意義を説明する．その後，本論文の構成について述べる．

1.1　本研究の背景

現在，我が国は様々な社会課題を抱えている．その中でも，ホームレス，放置自転車，ごみのポイ捨て問題など都市の特性や人の行動に関係の深い社会課題（都市問題）に関しては，自治体ごとに対策に向けた議論や施策検討が行われている．近年，こうした社会課題および都市問題のデータ利活用による課題解決アプローチが注目されており，2012 年の「電子行政オープンデータ戦略」策定[1]，2013 年に行われた G8 サミットでの「オープンデータ憲章[2]」への合意以降，政府を始め各自治体でもオープンデータ化の取り組みが普及している．大阪市ではオープンデータポータルサイト[3]で約 10,000 件のデータセットを公開しており，オープンデータを活用した課題解決に役立つアプリケーションの事例を紹介している．横浜市では，NPO 法人横浜コミュニティデザイン・ラボにより，横浜市のオープンデータを活用した地域課題解決のビジネス創出を支援するクラウドファンディングサイトが運営されている[4]．他にも京都市，さいたま市，新宿区など，各市区町村で市の統計情報や会議録など様々なデータの公開が進んでおり，オープンデータを活用した社会課題ないし都市問解決が各地で注目されている．さらに，2016 年に「官民データ活用推進基本法[5]」が公布されたことにより，より一層データ活用による課題解決が求められている．しかし，都市問題そのものの詳細なデータに関しては，公開だけでなくデータの収集もできていな

[1]https://www.kantei.go.jp/jp/singi/it2/pdf/120704_siryou2.pdf
[2]https://www.mofa.go.jp/mofaj/gaiko/page23_000044.html
[3]https://data.city.osaka.lg.jp/
[4]http://yokohama.localgood.jp/
[5]https://www.kantei.go.jp/jp/singi/it2/hourei/detakatsuyo_honbun.html

い状態である．例えば放置自転車に関して言えば，自治体によっては年単位かつ広域な統計データは存在するものの，このような粒度の粗いデータをもとに，具体的な原因特定や解決施策の実施に向けた分析を行うことは困難である．また，ポイ捨て問題に至っては自治体がデータを収集することができておらず，一部自治体が企業に調査委託した結果を保持する程度であり，オープンデータとしては公開されていない．

さらに，都市問題は様々な原因が複雑に絡み合っているため，解決施策実行に先立って自治体の抱える各種都市問題の原因や影響，都市問題同士の関係性を俯瞰して把握することが望ましい．しかしながら，このような都市問題間の関係が整理されたデータセットは存在しておらず，自治体によっては担当地域において生じている都市問題を把握できていないこともある．

1.2　本研究の目的と意義

本研究では前述の背景を受けて，都市問題解決に向けたデータ分析活用基盤を構築し，分析活用例を示すことで問題解決の一助となることを目的とする．特に，本研究では都市問題の中でも放置自転車問題や路上のポイ捨て問題など，都市の地理空間的な特徴と人の行動に関連の深い問題を対象とする．そのような都市問題のデータ利活用による課題解決において，(1) 都市問題そのものの詳細なデータが不足していること，(2) 都市問題間の因果関係が整理されていないことの２つの問題を解決するために，データ分析活用基盤の実体として (1) と (2) のそれぞれの問題解決に対応する，知識をグラフ状に整理したデータベース（ナレッジグラフ）を設計・構築する．また，ナレッジグラフ構築の過程で生じる欠損を，セマンティック Web 技術や機械学習などを用いて推定し補完する．さらに，都市問題の解決に向けて構築したナレッジグラフからの検索や可視化を通じた分析例を示す．これらのナレッジグラフ設計・構築・欠損推定補完・分析活用の手法群を同様の都市問題においてもある程度流用が可能となるよう留意し，一つの枠組みとして提案する．本研究の取り組み範囲を図 1.1 に，より詳細化した研究の全体図を図 1.2 示す．

ナレッジグラフとは，ある事象に関する知識を頂点と辺の集合で構成されるグラフにしたものであり，知識はグラフにおける頂点間の関係として記述される．すなわち，知識は主語（頂点）・述語（辺）・目的語（頂点）の三項構造（トリプル）として表現できる．本研

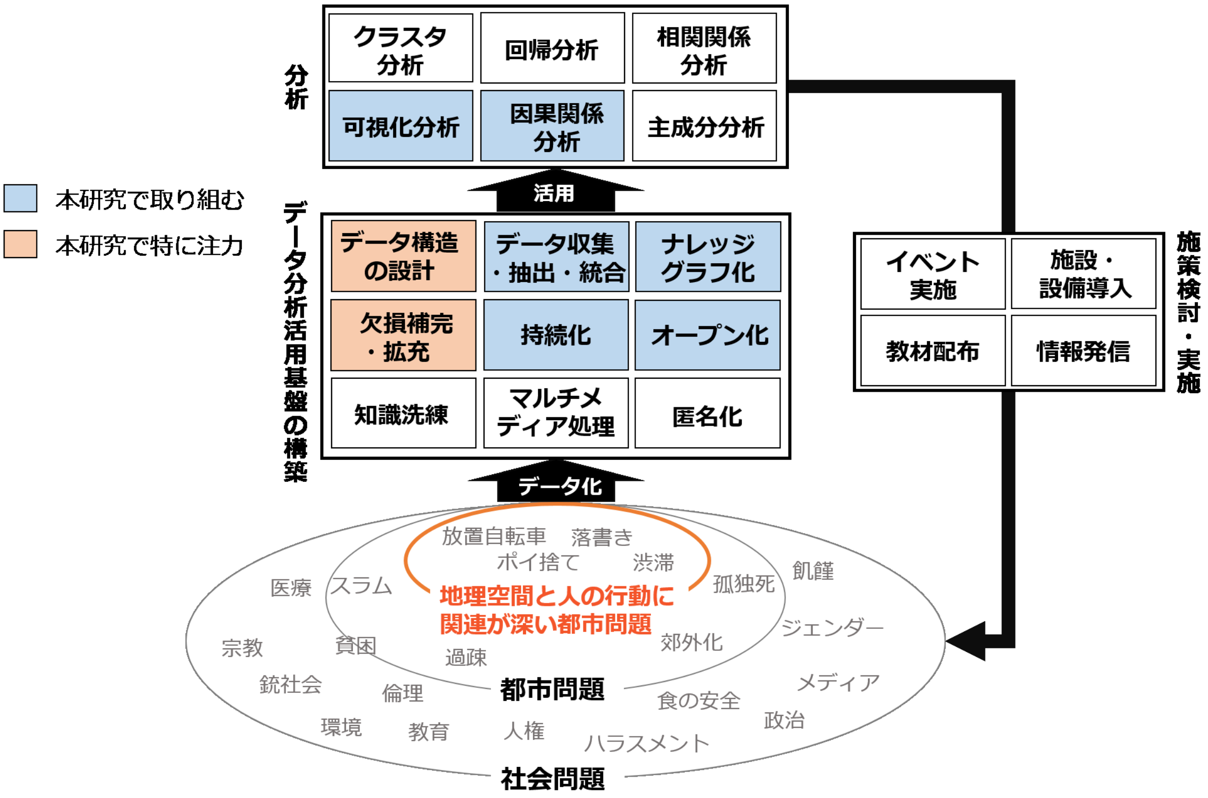

図 1.1: 本研究の取り組み範囲

究では各都市問題の時間的・空間的に詳細な情報や，都市問題間の原因や影響についての情報を記述したナレッジグラフを構築することで，データ分析活用基盤の実現を目指す．また，本研究ではナレッジグラフのデータ記述・管理・共有方法として Linked Data[6] を採用する．Linked Data はすべてのエンティティおよびリレーションに一意な Uniform Resource Identifier (URI) を付与することで，ナレッジグラフを統一的な形式で記述する．これにより，分野の異なる様々なデータの統一的な管理と高い拡張性を有し，データ同士のハイパーリンクを辿った柔軟な検索を可能にする．また，Linked Data は Web 標準化団体 World Wide Web Consortium (W3C)[7] により推奨されており，国内では総務省，内閣府等によりオープンデータの共有方法として推奨されている．

まず，本研究では都市問題そのものの詳細なデータが不足しており分析が困難である課

[6] https://www.w3.org/DesignIssues/LinkedData
[7] https://www.w3.org/

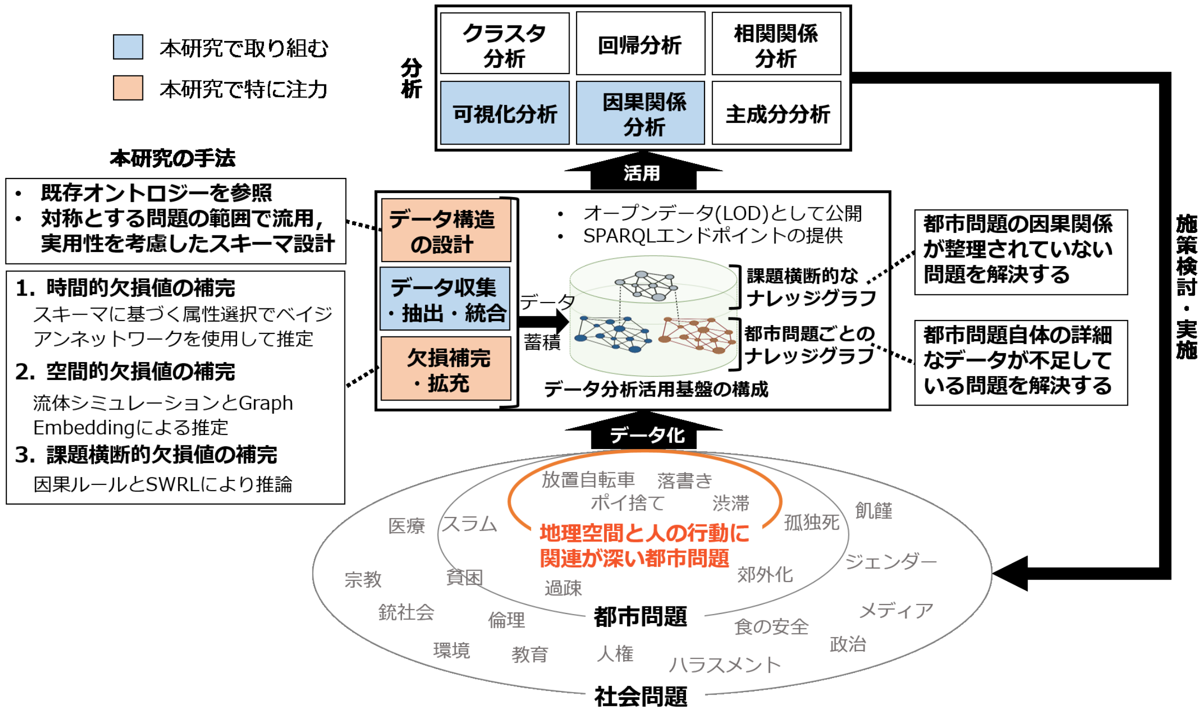

図 1.2: 本研究の全体図

題の解決に向けて，都市問題ごとにデータを収集して時間的・空間的に詳細な検索が可能なナレッジグラフを構築する．ナレッジグラフ化を行う上で重要となるデータ構造（スキーマ）を設計し，次にソーシャルセンサを活用してデータを収集し，設計したスキーマに基づいて Linked Data 化を行う．法律とコストの問題から街中にカメラやセンサなどを多数設置することは困難であるため，本研究では人とスマートフォンによるソーシャルセンサを用いてデータの収集を試みる．しかし，ソーシャルセンサにより集められるデータには時間的または空間的な欠損が生じるという問題がある．時間的・空間的欠損のあるスパースな状態のナレッジグラフでは，都市問題の可視化等の分析を行う際に充分な検索ができない．そこで，本研究では都市問題に関するナレッジグラフの時間的・空間的な欠損値について，機械学習的アプローチにより推定し補完する手法を提案する．ナレッジグラフの欠損補完については，近年，Translation-based モデルの TransE [9] の出現を機に Knowledge Graph Completion タスクとして注目されている．本研究で行う欠損補完についても Knowledge Graph Completion タスクに大別されるが，都市の地理空間的な特性と人の行動に関係の

深い都市問題を対象として，Translation-based モデルよりも対象とする問題に適したアプローチを用いて高精度に欠損補完を行う．具体的には，時間的な欠損値についてはベイジアンネットワークを使用し，空間的な欠損値については数値流体力学とニューラルネットワークを使用する新たな手法を開発した．

次に，複数の都市問題間の原因や影響に関するデータを収集してナレッジグラフ化する．まずスキーマを設計し，次に Web から収集した記事，SNS 情報，自治体オープンデータ，学術情報などを基に，自然言語処理とクラウドソーシングを用いて都市問題の因果関係や階層関係を抽出する．設計したスキーマに基づいて，抽出した因果関係や階層関係を記述するナレッジグラフを構築する．これにより，例えば都市問題の因果関係や階層関係のリンクをたどって課題の影響範囲を予測することや，悪循環および悪循環を引き起こす根源的課題を発見することが可能になり，自治体や NPO 法人などの課題解決実務者の一助になると考える．しかし，収集したデータから直接的に記述された因果関係を抽出するため，異なる文脈で出現したエンティティ同士の因果関係については欠損が生じていることがある．そこで，この因果的な欠損をセマンティック Web ルール言語により推論して補完する．ここで，本稿で述べる “因果関係” とは，単にある事物の要因や影響として指摘されうる関係であり，その他の科学的な意味合いを指していないという点に注意されたい．

最後に，本研究で構築したナレッジグラフを検索エンドポイントとともに Linked Open Data (LOD) として公開することで，第三者の利活用を可能にする．さらに，ナレッジグラフの活用例を示し，都市問題解決の一助としての有用性を考察する．都市問題の時空間的に詳細なナレッジグラフの活用に関しては，放置自転車を例として時空間的な台数変化をアニメーション表示するアプリケーションを開発した．これにより，都市の現状把握と市民の問題意識向上やさらなるデータ収集が期待される．さらに東京都と連携することで，開発したアプリケーションを青少年・治安対策本部のホームページを通して紹介した．都市問題の因果関係に関するナレッジグラフの活用に関しては，都市問題の因果関係から構成される悪循環および，複数の悪循環を引き起こす根源的な課題を発見する検索を行った．さらに，大阪市市民局協力のもとで，都市問題の専門家との議論の場において検索結果を提示して評価を行った．

したがって，本研究の意義は次の通りである．

1. 都市問題のデータ利活用による課題解決に向けた，ナレッジグラフの設計・構築・公開・欠損推定補完・分析活用に関する一連の手法からなる枠組みの提案

2. 自治体との連携による研究成果の社会実装に向けての評価

特に，本研究における一番の主張点はナレッジグラフの欠損推定補間手法の提案である．

本論文では，以降 2 章で要素技術および関連研究について説明し，3 章で都市問題ごとのナレッジグラフの設計と構築手法について述べる．4 章で 3 章で構築するナレッジグラフの時間的・空間的な欠損値の推定手法と，ナレッジグラフの分析例について述べる．5 章で都市問題の因果関係と自治体事業予算情報を含むナレッジグラフの設計と構築，および分析手法について述べる．最後に 6 章でまとめとする．

第2章　関連研究

本章では，本論文で構築するナレッジグラフに関連した研究について説明する．まずはナレッジグラフと，本研究で扱うナレッジグラフの形式として関連の深いLinked Dataおよびオントロジーに関する概要を説明する．次に社会課題および都市問題に関連するナレッジグラフの研究について説明する．そして，自然言語処理や機械学習，クラウドソーシングを用いたナレッジグラフの構築に関する説明をする．さらに，機械学習やルール推論によるナレッジグラフの欠損補完や予測に関する研究について説明する．

2.1　ナレッジグラフとLinked Dataの概要

本節では，ナレッジグラフと，本研究でナレッジグラフとして扱うLinked Dataおよびオントロジーについて説明する．

2.1.1　ナレッジグラフ

意味ネットワーク (Semantic Networks) [58] や概念グラフ (Conceptual Graph) [57] など，従来から知識表現の手法としてグラフ，ネットワークが用いられている．ナレッジグラフとは，ある事象に関する知識をグラフ理論におけるグラフで表現したものであり，知識はグラフにおける頂点と辺を用いた関係として記述される．2012年にGoogleがナレッジグラフを用いて検索エンジン結果を拡張するサービスKnowledge Graph[1]を提供したこともあり，近年ナレッジグラフへの注目が高まっている．米国IT調査企業のGartnerが公開した2018年版先進テクノロジのハイプ・サイクル[2]によると，ナレッジグラフはInnovation

[1] https://www.google.com/intl/bn/insidesearch/features/search/knowledge.html

[2] https://www.gartner.com/en/newsroom/press-releases/2018-08-20-gartner-identifies-five-emerging-technology-trends-that-will-blur-the-lines-between-human-and-machine

Triggerの時期として注目されており，過度な期待を寄せられるピーク期の前として位置づけられている．ナレッジグラフと同じ時期に位置づけられているものとしては，汎用AI，量子コンピューティング，ブロックチェーンによるデータ・セキュリティ等がある．このように現在ナレッジグラフが注目されているのは，ビックデータ時代における種々雑多なデータを管理し，有用な知識を意味的に検索する需要が高まっていることが背景として考えられる．特に，WikiData[3] [65]，Freebase[4] [8]，YAGO [60][27][50] などの大規模知識ベースを始め，様々な分野のデータがLinked Dataの形式のオープンなナレッジグラフとして提供され始めたことが，近年のナレッジグラフの注目と研究の進歩に貢献している．

2.1.2 Linked Data

Linked Dataとはデータを機械処理が容易な形式でWorld Wide 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．特にオープンライセンスで公開されたLinked DataはLinked Open Data (LOD)と呼ばれる．Linked Dataとは，Webの発明者ティム・バーナーズ=リーの定義した以下の基本原則[5]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

Resource Description Framework (RDF) とはWeb標準化団体W3Cが勧告してい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Uniform Resource Identifier (URI) の形式で記述される．ま

[3]https://www.wikidata.org/wiki/Wikidata:Main_Page
[4]https://developers.google.com/freebase/
[5]https://www.w3.org/DesignIssues/LinkedData.html

た，N-Triples，RDF/XML，JSON-LD，Terse RDF Triple Language (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 Protocol And RDF Query Language (SPARQL)[6]は RDF に対して問い合わせが可能なクエリ言語である．Linked Data はナレッジグラフの一形式であり，統一形式でのデータ共有と SPARQL による柔軟な検索を容易にしたことにより，近年のナレッジグラフ普及に大きく貢献している．実際，大規模知識ベースとして認知されている WikiData や Freebase を始め，多くのナレッジグラフは Linked Data の形式を取っている．

図 2.1 に示す Linked Open Data cloud diagram[7]から分かるように，行政，地理，ライフサイエンス，言語，メディア，書誌，SNS，ユーザ生成コンテンツ，クロスドメイン等，様々な分野のデータが LOD 化されている．この図には規模が 1000 トリプル以上であり，図内の他の LOD とのリンクを 50 以上含むデータセットを条件に本サイトが認知しているものを図示しているが，この条件を満たしていてもこの図内に含まれていない LOD は多数存在している．

2.1.3 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 [73]．人工知能の立場からは「概念化の明示的な規約[8]」と定義されている [73]．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，Suggested Upper Merged Ontology (SUMO) [47]， Descriptive Ontology for Linguistic and Cognitive Engineering

[6]SPARQL は再帰的頭字語
[7]`http://lod-cloud.net/`
[8]`http://www-ksl.stanford.edu/kst/what-is-an-ontology.html`

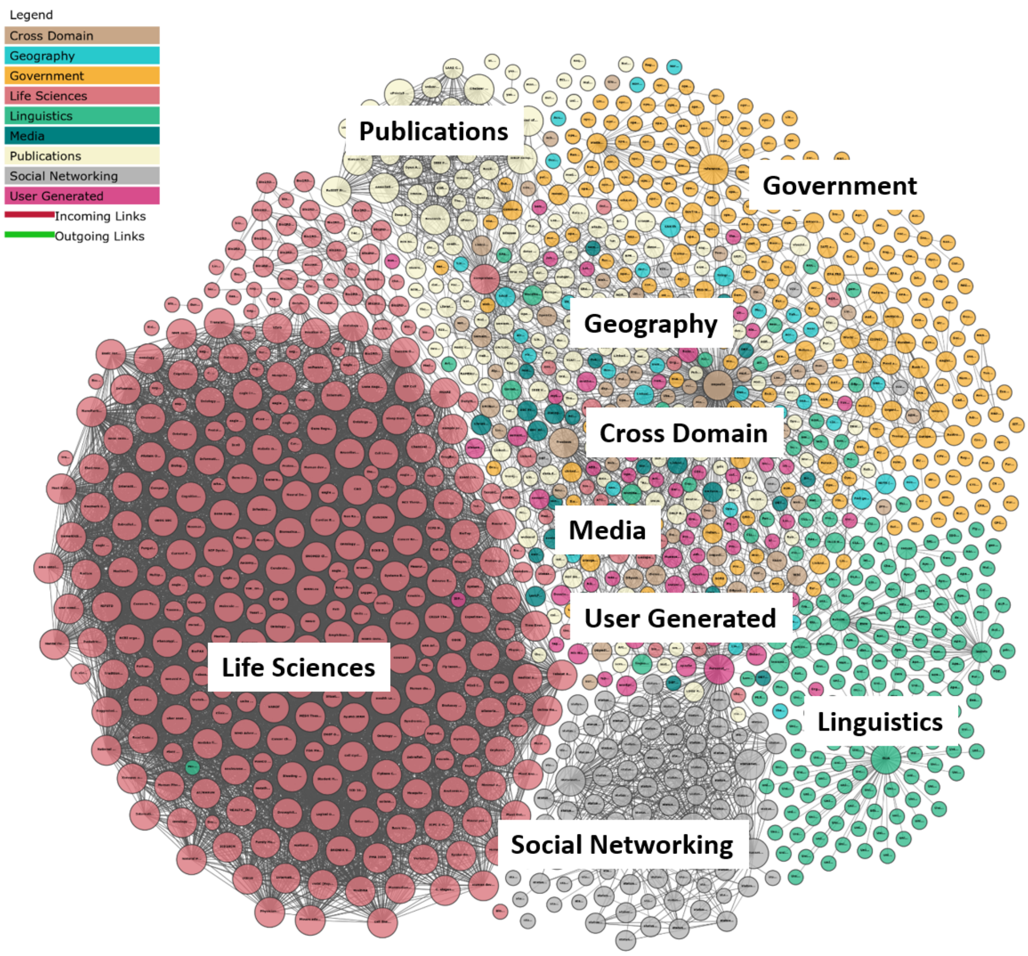

図 2.1: Linked Open Data cloud diagram (2017 年 8 月時点)

(DOLCE) [21] や Yet Another More Advanced Top-level Ontology (YAMATO) [44] 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 [1] や医療オントロジー [76] 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 [77] が挙げられる．オントロジーの RDF 拡張としては Web Ontology Language (OWL)[9]が存在し，現在多くのオントロジーが OWL を使用して記述されている．OWL は 2004 年に W3C 勧告となり，2009 年には OWL 2 が W3C 勧告となっ

[9]https://www.w3.org/TR/owl-guide/

ている．本研究ではナレッジグラフと言った場合には，主に Linked Data や OWL オントロジーなどを指す．

2.2　社会課題および都市分析に関するナレッジグラフの研究

これまでに，社会課題解決に向けた幾つかのナレッジグラフが構築されているため，本節で説明する．Szekely らは人身売買問題解決に向けて，Web の悪質広告から収集した情報を基にナレッジグラフを構築した．この研究では悪質広告に記載されている電話番号や人物プロフィールだけでなく，人物画像の類似度を算出しナレッジグラフに記述している．また，構築したナレッジグラフを活用した迷子検索システムを開発し，法執行機関や NGO に導入している [61].

Deng ら [16] は野生動物のロードキル事故における，ソーシャルセンサによるデータ収集および管理のための Linked Data を提案している．この研究では Facebook を用いたクラウドソーシングによりロードキル情報を収集し，既存語彙として Semantic Sensor Network Ontology[10] [12]，Friend Of A Friend (FOAF)[11]，Semantically Interlinked Online Communities (SIOC) core ontology[12]を使用して Linked Data 化している．

Shiramatsu ら [55] は社会課題の解決目標を LOD 化するためのデータモデルを提案している．さらにこの LOD 活用し，課題解決者同士のマッチングをすることでシビックテックを促進させる手法を提案している． Web アプリケーションとして Goalshare を開発しており，実際に社会課題解決に向けてシビックテックイベントでの活用を行っている．この研究で構築された LOD は課題解決目標に関するものであり，都市問題そのもののデータや因果関係については記述されていないが，将来的に本論文で構築したナレッジグラフとのリンクにより課題解決促進が期待できる．

都市データのナレッジグラフ化の取り組みとしては Lopez らの研究 [40] が挙げられる．この研究では，統計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．センサーからストリームデータを取得し，IBM InfoSphere

[10] https://www.w3.org/TR/vocab-ssn/
[11] http://xmlns.com/foaf/spec/
[12] http://sioc-project.org/

Streams[13]と C-SPARQL [4] 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都市問題のデータに関しては，同様の取り組みを行うことは難しい．

Santos ら [53] は市の様々な指標の値を算出するために Linked Data を構築し，可視化をサポートする quality of experience (QoE) Indicators Ontology を新たに提案している．また，Linked Data からデータを取得して可視化ウィジェットを生成できるダッシュボードを開発している．

Pileggi ら [49] は首都圏の動的かつ詳細な指標を表すために，Conceptual Framework を定義し，OWL-DL オントロジーとして実装している．データ構造の理解を容易にするために，Indicator, Data, Profiling, Computations, Geographic Context の 5 つのサブオントロジーに分割している．

OpenSpending.org では各国の自治体が公開する予算・支出情報のオープンデータのカタログサイトである．現在 75 カ国のデータが集約されている．これらのデータを LOD 化したものに LinkedSpending [28] がある．統計データの RDF データモデルを表現する RDF Data Cube Vocabulary[15]に従って OpenSpending のデータを Linked Data 化している．これらの Linked Data には都市問題情報は含まれていないが，本研究で構築するナレッジグラフとリンク出来る可能性がある．

2.3　ナレッジグラフの構築に関する研究

本節では，本研究で使用する自然言語処理とクラウドソーシングを用いたナレッジグラフの構築に関する研究を説明する．

Demartini ら [15] はクラウドソーシングを用いた Linked Data のための Entitiy Linking 手法を提案している．クラウドソーシングはリンクの品質を向上させるために使用され，信頼性の低いユーザを排除し，一貫性のない結果を統合するための確率ネットワークを開発

[13]http://www-03.ibm.com/software/products/ja/ibm-streams
[14]http://www.dublinked.ie/
[15]https://www.w3.org/TR/vocab-data-cube/

している．Celino ら [11][10] は，Linked Data における Point of interest (POI) データと写真をリンクするためのモバイルアプリケーションを開発している．ユーザにインセンティブを与えるために Game with a purpose (GWAP)[64] の概念を導入し，クラウドソーシングによりデータをリンクさせている．しかしながら，これらの手法は正解データが存在することを前提としており，都市問題の因果関係のように正解データが存在しない問題に対して適用することは困難である．

Nguyen ら [46] は Conditional Random Fields を使用し Twitter からユーザの行動を推定し，Linked Data を構築する手法を提案している．ユーザの行動や日時，場所，目的等をトリプルとして構築している．主に，震災時の行動分析や，ナビゲーションに応用が可能である． LODfier[3] は Wikifier[42] を用いた Named Entity-Recognition により非構造化テキストからエンティティを抽出し，DBpedia [2] や WordNet にリンクする RDF を生成する．Linked Data 構築に向けた自然言語処理手法は数多くあるが，我々の知る限り，これらの単語抽出手法を用いて因果的な単語を完全自動で抽出することは，精度の面で十分でない．したがって，自然言語処理とクラウドソーシングを組み合わせることで，都市問題の因果関係を半自動的に抽出する手法を開発した．

2.4　ナレッジグラフの欠損補完と機械学習への適用

本研究と同様に社会データに関するナレッジグラフを補完・推定する取り組みとしては Bischof ら [6] の研究が挙げられる．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，Eurostat[16]，United Nations Statistics Division(UNSD)[17]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づいて 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リング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

[16]http://ec.europa.eu/eurostat
[17]http://unstats.un.org/unsd/default.htm

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，定期的なクローリング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．

グラフデータを元に機械学習を行うために，グラフ内のエンティティをベクトル空間に埋め込む手法が存在する．DeepWalk [48] はグラフ上でランダムウォークを行うことで頂点のシーケンスを生成し，word2vec [41] により頂点のベクトル表現を学習することで注目を集めた．LINE [62] は DeepWalk の改良版として提案され，DeepWalk が深さ優先探索であるのに対し，幅優先探索で頂点のベクトル表現を獲得する．さらに，これら二つの改良手法として node2vec [22] が提案され，遷移確率に二つのパラメータを導入することで，深さ優先探索と幅優先探索を調整することが可能である．

DeepWalk，LINE，node2vec はエッジがセマンティクスを持たないグラフに対する手法であるが，RDF2vec [51] はエッジがセマンティクスを持つ RDF を対象に，エンティティをベクトル空間に埋め込むことができる手法として提案された．RDF2Vec はまず RDF 上で指定した深さで Graph Walk を行い，頂点 → エッジ → 頂点 → エッジ... の walk パスを生成する．次に，Weisfeiler-Lehman Graph Kernel [54] を RDF に対して行う Weisfeiler Lehman Subtree RDF Graph Kernel [14] を用いて各頂点のサブグラフに関して頂点とエッジのラベル更新（リラベル）を行う．リラベルグラフ上で walk パスに基づく Graph walk を行うことで，エンティティに関するシーケンス集合を得る．これを元に word2vec を行うことで，エンティティのベクトル表現を得る．Kejriwal ら [33] は RDF2Vec がクロスドメインデータにはうまく作用するが，GeoNames[18]等の緯度経度リテラルが重要な地理データにはうまく適用できない点に着目した改良手法を提案している．GeoNames から頂点間のエッジの重みが測地線距離の近似になるようにグラフを構築し，DeepWalk を用いて頂点のコサイン類似度が空間的近接度に近似するように超球面に埋め込む．その際に，ランダムウォークが局所的に完結しないように減衰重みを導入している．

ナレッジグラフの欠損補完および Link prediction には Translation-based model が近年よ

[18]http://www.geonames.org/

く使用される．最初の Translation-based model の TransE [9] はトリプル $(head, relation, tail)$ において，$head + relation - tail = 0$ に近づけるように頂点とエッジをベクトル表現として学習する．TransH [66] や TransR [39] は $relation$ を別の空間に埋め込むことで，同一エンティティからの複数プロパティに対応している．

薬学分野においては，薬物相互作用 (Drug-Drug Interaction, DDI) の発見および薬物データベースの欠損補完のために，ナレッジグラフを用いたルール推論を行う研究が進んでいる．Moitra ら [45] は薬物動態学の DDI を Semantic Application Design Language [13] を使用してモデル化し，いくつかの酵素に関する相互作用を SWI-Prolog[19] を使用して推定している．Herrero-Zazo ら [25] は薬物動態学と薬力学の間の相互作用のための包括的なオントロジーとして DINTO を提供し，DINTO と Semantic Web Rule Language (SWRL)[20] を使用して "may interact with" のような起こりうる可能性のある作用を推定する研究を行っている．本研究では，これらの DDI におけるナレッジグラフを用いた推論の研究を参考にし，都市問題の因果リンクの補完と悪循環の発見および評価を行った．

[19] http://www.swi-prolog.org/
[20] https://www.w3.org/Submission/SWRL/

第3章　都市問題ごとのナレッジグラフの設計と構築

本章では，図 3.1 に示す都市問題のデータ分析活用基盤のイメージ図における，都市問題ごとのナレッジグラフの設計と構築について説明する．

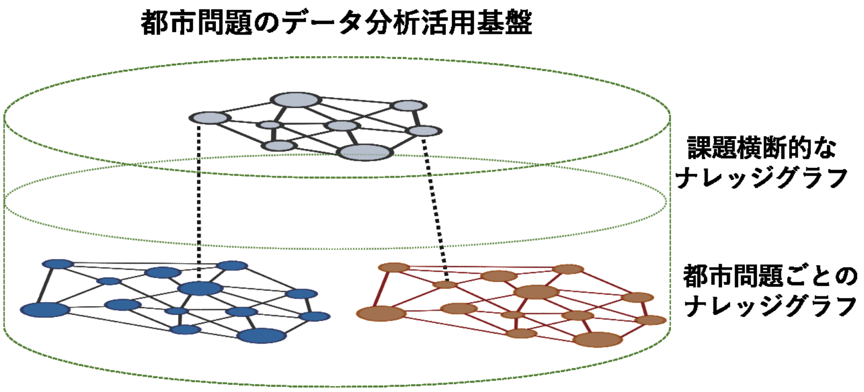

図 3.1: 都市問題のデータ分析活用基盤のイメージ図

3.1　都市問題に関するナレッジグラフのスキーマの設計

本研究では Linked Data の形式でナレッジグラフを構築する．本節では特定の都市問題の時間的・空間的に柔軟な分析が可能な Linked Data 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手法について説明する．Linked Data 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担保しつつ，都市問題解決に向けた検索・分

析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都市問題に関するスキーマ設計の方法論として示す．

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．我々の知る限り，都市の地理空間的な特徴と人の行動に関連の深い問題において，同様の問題への適用可能性やデータの分析活用を考慮して構築された既存オントロジーや，オントロジー構築の方法論などは存在していない．そこで本研究では，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 [43][75] を参考に，よりドメイン要件と分析活用に焦点を当て，同様の問題に適用できるように考慮して Linked Data 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，都市問題の一例として放置自転車問題を対象とし，スキーマを設計して Linked Data を構築する．本研究で提案する Linked Data 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. 対象とする都市問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. クラス・インスタンス・プロパティの設計
    a. 既存オントロジーを軸としてクラスを設計
    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設計

3.1.1 ドメイン要件の抽出

まず，対象とする都市問題を Linked Data 化する上で骨組みとなる既存オントロジーを選択する．Linked Data をオントロジーに基づいて構築することで，オントロジーの制約に基づく推論が可能になることに加えて，Linked Data 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 Linked Data をデータセットや API として活用す

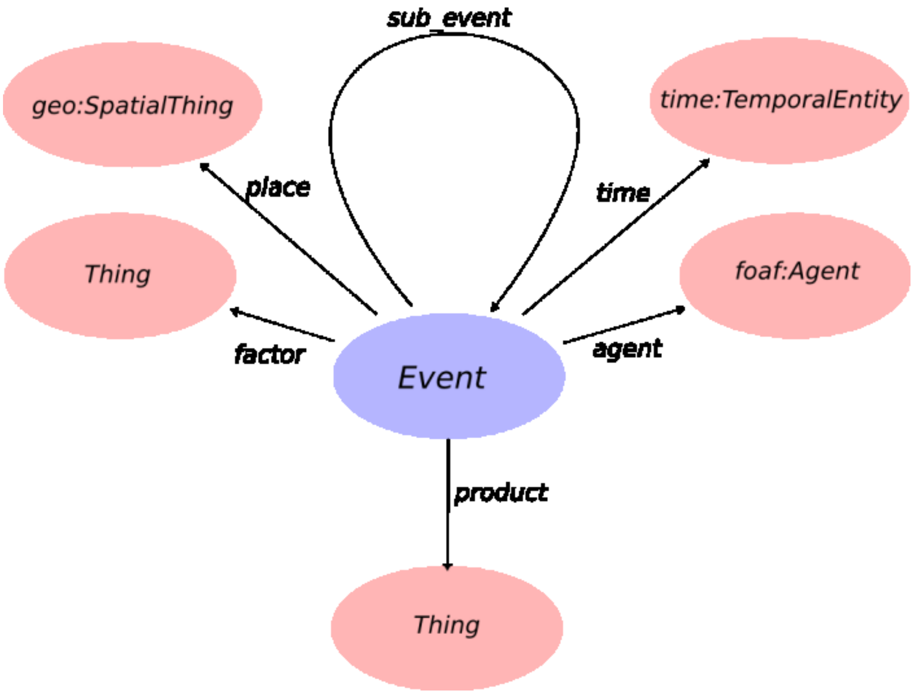

図 3.2: Event Ontology (http://purl.org/NET/c4dm/event.owl)

る際に利便性を向上させる．本研究では都市問題をイベントとして捉えて，図 3.2 の Event Ontology[1](以下，EO) をスキーマの骨組みとなるオントロジーとして選択した．

次に，都市問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手動で抽出する．放置自転車の場合では “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．さらに，オントロジーに相当するプロパティが定義されていないが， 人間の目で見て重要であると思われるキーワードについても抽出する．

抽出したキーワードを表 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`

表 3.1: 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

3.1.2　クラス・インスタンス・プロパティの設計

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，Linked Data 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の Linked Data（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．

$$\text{IllegallyParkedBicycles} \sqsubseteq \text{Event}$$

$$\text{IllegallyParkedBicycles} \sqsubseteq \exists\text{place.SpatialThing}$$

$$\text{IllegallyParkedBicycles} \sqsubseteq \exists\text{time.TemporalEntity}$$

$$\text{IllegallyParkedBicycles} \sqsubseteq \exists\text{weather.WeatherState}$$

$$\text{IllegallyParkedBicycles} \sqsubseteq \exists\text{factor.Thing}$$

$$\text{IllegallyParkedBicycles} \sqsubseteq \exists\text{agent.Agent}$$

$$\text{IllegallyParkedBicycles} \sqsubseteq \exists\text{product.Thing}$$

$$\text{IllegallyParkedBicycles} \sqsubseteq \exists \text{value.Integer}$$

IllegallyParkedBicycles クラスは放置自転車の集合を意味し，Event クラスのサブクラスとする．場所，時間，天気，人物，要因，引き起こす事象，放置自転車の台数を持つ．

次に表 3.1 を参考にインスタンスを設計する．図 3.3 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．表 3.1 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 Linked Data 化する際には，図 3.3 のインスタンスには特定の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，特定の観測地点における放置自転車が引き起こす事故情報なども取得することが困難であるため，event:factor 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 Linked Data 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 Linked Data スキーマを設計した．

3.2 都市問題の時空間的に詳細なデータの収集

本節では，都市問題の詳細なデータの収集方法を放置自転車を例に説明する．

本研究では，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者 (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

[2] https://www.lancers.jp

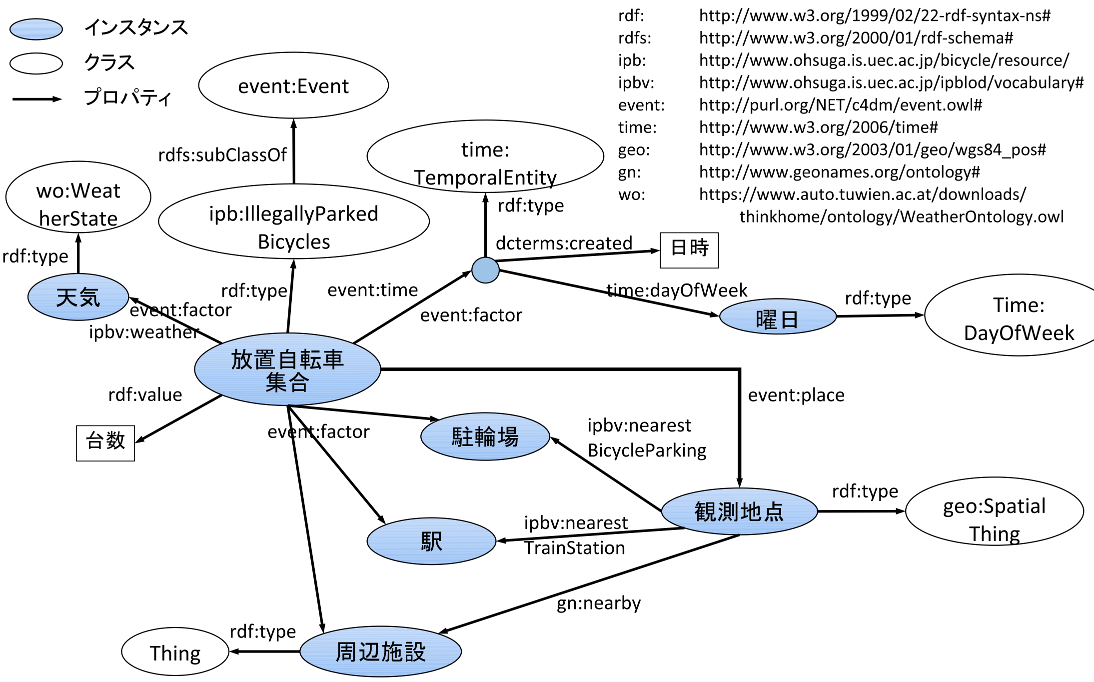

図 3.3: インスタンス加えて設計した LOD スキーマ

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．

しかしながら，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．一方，SNS 上でキーワード検索を行ったところ，放置自転車に問題意識を持った投稿が多く見られることがわかった．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

したデータ収集を行った．

本研究では，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 [29]．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．図 3.4 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図 3.4 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カが表示される．このマーカをドラッグすることで位置情報の修正が可能であるため，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 897 件であった．

さらに，3.1.1 節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]や 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 サイトから収集した．

3.3　ナレッジグラフの構築と公開

3.3.1　時系列データに基づくナレッジグラフのリアルタイム構築

3.1 節で述べた Linked Data スキーマに従い，3.2 節で収集した放置自転車に関する一次データを Linked Data 化する．Linked Data 化の手順を図 3.5 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付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検索する．放置自転車 LOD 内に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

[3]http://www.ohsuga.is.uec.ac.jp/bicycle/tweet_submit
[4]https://developers.google.com/places/?hl=ja
[5]https://developer.foursquare.com/

図 3.4: ツイート Web アプリの動作画面

め，最寄りの施設情報を Google Places API を使用して取得する．Google Places API で施設情報を取得できなかった場合は，Foursquare API を使用して取得する．最寄りの施設名に「前」や「北口」などの単語をつけたものを観測地点の名称とし，これを基に観測地点のリソースの URI を設計する．さらに，観測地点から半径 20 メートル以内の周辺施設を検索する．得られた周辺施設に URI を割り当ててリソースとして構築し，観測地点のリソースから gn:nearby プロパティでリンク付けする．Google Places API では施設のタイプ情報を得ることが可能であり，本研究ではこのタイプと LinkedGeoData [59] 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! リバースジオコーダ API[6] 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[7] 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．このように，ツイートを取得後に各種 API を用いて観測地点の情報を取得し，前述の Linked Data スキーマに

[6] http://developer.yahoo.co.jp/webapi/map/openlocalplatform/v1/reversegeocoder.html
[7] http://geonames.jp

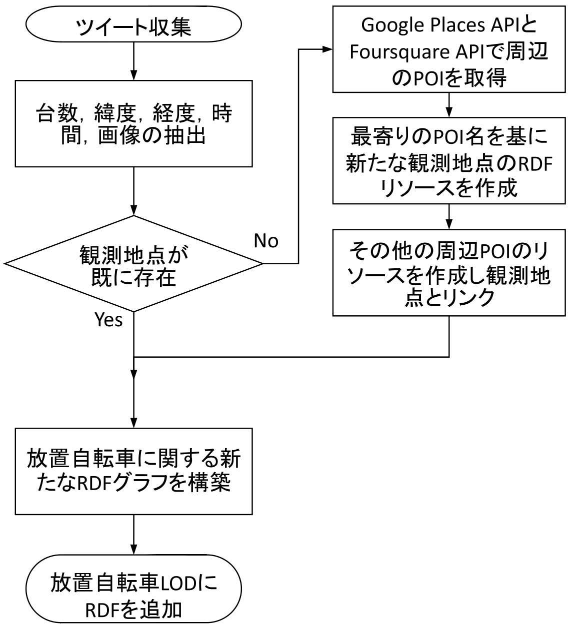

図 3.5: 一次データ収集と LOD 化の手順

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 3.6 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[8]を通して Web 上に公開している．

3.3.2 非時系列データに基づくナレッジグラフの構築

放置自転車問題以外にも，ゴミのポイ捨て問題，ホームレス問題，交通事故，交通渋滞，夜道の犯罪など，都市の地理空間的な特徴と人の行動に関係の深い都市問題は他にも存在する．そこで，このような都市問題に関してを本研究を横展開し，次の例としてゴミのポイ捨て問題に関する詳細なナレッジグラフの構築を行った．

路上でのゴミのポイ捨ては多くの都市が抱える問題である．ポイ捨てゴミは道路交通の阻害や景観破壊，環境破壊だけでなく，火災の原因になることもある．路上のポイ捨て状況をナレッジグラフ化することで，可視化や関連データとのつながりの検索を通して，これまで見えてこなかった新たな原因発見につながる可能性がある．これにより，都市部に

[8]http://www.ohsuga.is.uec.ac.jp/sparql

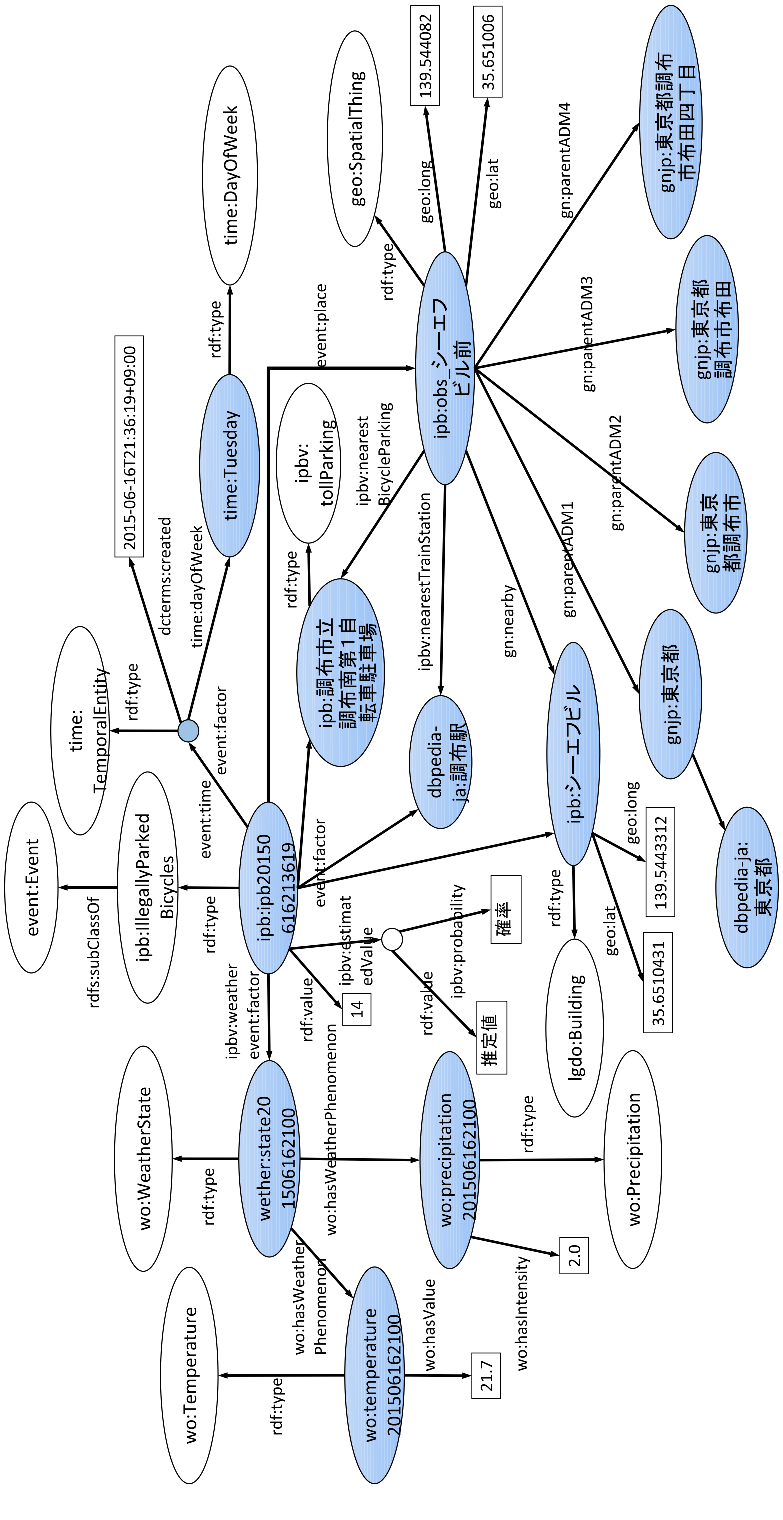

図 3.6: 実際に構築される放置自転車 LOD の一部

おけるゴミ箱や灰皿の設置や，清掃活動などの効果的な計画を行うことが可能になると考える．

本研究では，株式会社ピリカより「タカノメ[9]」で収集した公開可能なポイ捨てデータの提供を受けた．タカノメでは，自治体や議員から依頼を受けて，路上のポイ捨て調査を行い分析結果を提供する．ポイ捨て調査は駅周辺の道路を対象に行われ，スマートフォンアプリケーションによる動画撮影後に，画像認識と人手による修正によってデータを作成している．現場での撮影は一日のうちに行われ，ある一時点での街の状況を記録している．したがって時系列データではなく，一つのゴミごとに緯度経度とゴミの種別が付与されたデータになっている．本研究ではこのポイ捨て調査結果のデータを基にナレッジグラフを Linked Data 形式で構築・公開した．

ポイ捨て調査結果のナレッジグラフのスキーマに関しては，提供データということから 3.1 節で述べた方法論を完全に適用することは困難であった．したがって，企業側の利便性を損なわないように元データの構造を維持しつつ，3.1 節で述べた方法論の 1.(a),2.(a) を部分的に適用することでスキーマを設計した．なお，設計の際には企業側と密に議論を行い，データの共有と利活用の側面を最重視しつつ既存オントロジー制約に違反しない構造とした．元データは放置自転車と同じくスマートフォンで取得したデータであるが，調査方法や元データの構造がセンサデータとしての意味合いが大きかったことから，センサデータのオントロジーである Semantic Sensor Network Ontology (SSN) [12] における SOSA Ontology [30] を基に記述している．その他，必要なプロパティおよびクラスを独自に拡張している．図 3.7 に実際に構築されるポイ捨て調査結果の LOD(ポイ捨て LOD) の一部を示す．

各ポイ捨てゴミごとに緯度経度情報とゴミの種類が付加されており，ポイ捨てゴミは `sosa:Observation` クラスのインスタンスの値としてリンクされている．`sosa:Observation` クラスのインスタンスは一つの観測データに対応しており，スマートフォンアプリケーションで撮影した動画を意味している．したがって，`geo:location` プロパティの値として `geo:SpatialThing` クラスのインスタンスを持ち，このインスタンスは住所情報に加えてジオメトリクラス `grq:Geometry` のインスタンスを持つ．ジオメトリクラスのインスタンスは，実際にソーシャルセンサが観測した経路の始点と終点の値をポリラインとして持

[9] `https://research.pirika.org/`

ち，その距離を schema:distance プロパティの値として持つ．このように設計することで，GeoSPARQL [5] による地理空間的な検索が可能になる．ゴミの種類と分類構造は企業側の元データセットとの互換性を意識しているため，本来オントロジーとしては好ましくない分類構造も含まれている．例えば，分類においてタバコの本数のような数を基準とした分け方と，プラスチック (色付き) とプラスチック (透明) のような色を基準とした分け方が混在している．この点について，今後分類の基準を統一化し，かつ検索の際にこれまでと同じ結果を返すことのできるように改善することを検討している．

構築したポイ捨て LOD は株式会社ピリカのサイト上[10]で公開している．

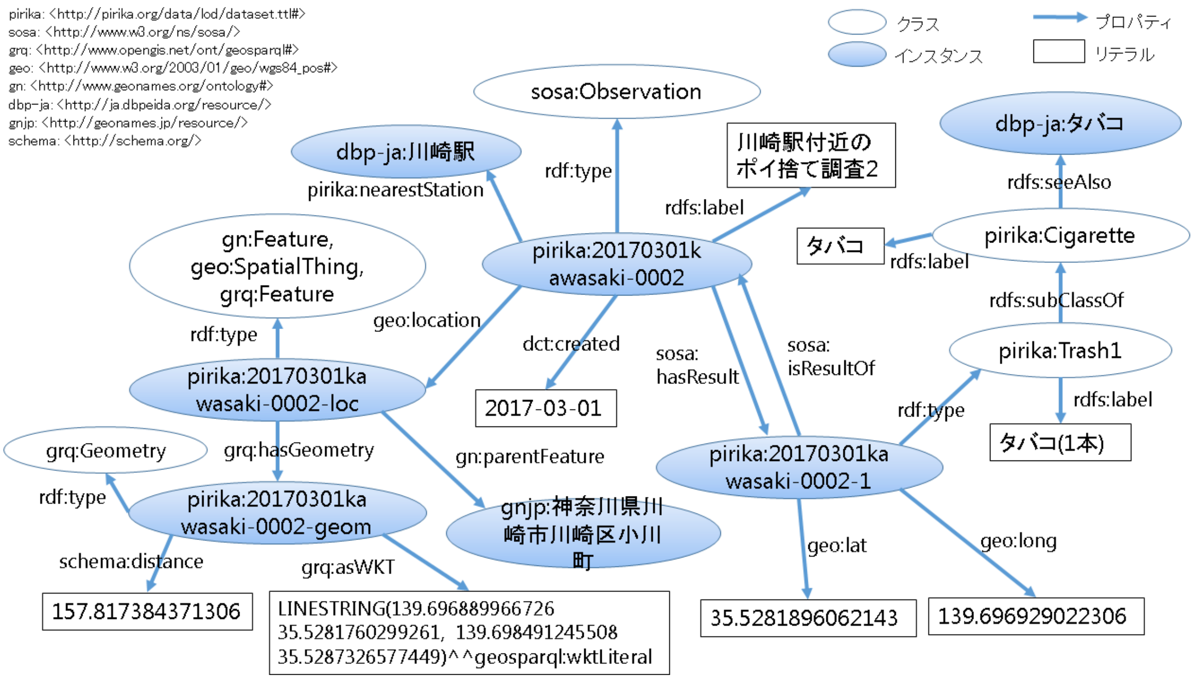

図 3.7: 実際に構築されるポイ捨て調査結果の LOD の一部

[10]https://www.pirika.org/data/lod/dataset.ttl

第4章　都市問題ごとのナレッジグラフの時空間的な拡充と分析

本研究で対象とする都市問題のデータは人手およびソーシャルに収集されるため必ず欠損が生じる．本研究ではナレッジグラフを人手で設計することに加えて，ナレッジグラフの欠損を推定・補完することにより，都市問題ごとのナレッジグラフを半自動的に拡充し続ける手法を提案する．まず，放置自転車 LOD について時間的な欠損値をベイジアンネットワークを用いて推定する手法について説明し，次に，数値流体力学シミュレーションを使用した放置自転車 LOD の空間的な欠損値の推定手法について説明する．さらに，Graph Embedding 手法とニューラルネットワークを用いて，空間的な欠損値の推定手法を拡張し，ポイ捨て調査結果の LOD について適用した実験について説明する．最後に，時間的または空間的に拡充された LOD を，検索や可視化によって分析した例について説明する．

4.1　ベイジアンネットワークを用いた時間的な欠損値の推定

放置自転車問題において放置台数の観測は人により行われるため，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．そこで，本研究では，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 3.1 から，放置自転車は場所や時間など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たためである．

4.1.1　欠損属性値の補完

まず，3.3.1 節で構築した 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リテラル値と仮定し，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，降水の有無，駐輪場料金の有無，周辺施設のタイプとした．ここで，周辺施設のタイプは表 3.1 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 でタイプを取得することができなかったため，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 14 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 $Hour = \{1, ..., 24\}$，降水の有無 $Rain = \{0, 1\}$，駐輪場料金の有無 $Fee = \{0, 1\}$，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 [31] を用いて 4 つに分類し，$Num = \{1, ..., 4\}$ とする．台数の範囲は，0〜6，7〜17，18〜35，36〜100 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，Fee を除いた観測データ o_1, o_2 の類似度は式 4.1 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．

$$sim(o_1, o_2) = \frac{|o_1 \cap o_2|}{|o_1 \cup o_2|} \tag{4.1}$$

4.1.2　ベイジアンネットワークによる放置自転車台数の推定

4.1.1 節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，4.1 節の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ライブラリとして Weka [23] を使用した．入力データは要素数 16 のベクトル

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

の集合 O であり，観測データ数は 897 件である．したがって，897 件のデータを訓練データとして，既知の観測地点のみ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．

4.1.3 推定精度評価を通じたパラメータ設定

収集したデータ数が813件の段階では，ベイジアンネットワークのパラメータにおいて，探索アルゴリズムとして HillCimber を選択し，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 としたときに最も制度が高くなり，推定精度は 65.2%となった．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，スキーマ設計の際に除外した項目について着目した．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る．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[1]の API から取得した．尚，放置自転車 LOD のスキーマとして地理情報を詳細に記述しているため，その情報を頼りに取得可能である．提供されている統計情報の内，各市町村における「可住地面積 $1km^2$ 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サイド 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．

[1] `http://www.e-stat.go.jp/`

次に，周辺施設のタイプに着目した．3.3.1 節で述べたナレッジグラフ構築の際には，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 3.1 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設の元に行った．しかし，実際には放置自転車の要因として，これらの周辺施設以外も要因になりうるという新たな仮設を立て，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みられるため，周辺施設のタイプをクラスタリングすることで対応した．放置自転車 LOD では，周辺施設のタイプを LinkedGeoData 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` しか無いタイプに関しては現タイプを使用するものとした．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
```

したがって，入力データは要素数 56 のベクトル

$$o_i \in Location \times Day \times Hour \times Precipitation$$
$$\times Temperature \times DailyFee \times MonthlyFee$$
$$\times Density \times Commuters \times Pharmacy$$
$$\times \cdots \times PublicTransportThing \times Number$$

の集合 O である．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に全体から 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，シード値を変えながら 10 分割交差検証を計 10 回行ったときの平均推定精度が 70.3%と向上した．さらに，継続的なデータ収集を行いデータ数が 897 件まで増加した段階では，探索アルゴリズムとして HillClimber に，最大親ノード数を 2 に変更したときに平均推定精度が 70.9%まで向上し，本実験において最も高い結果を得た．詳細な結果は 4.1.4 節で記す．

4.1.4 評価

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 897 件であった．表 4.1 に観測データの統計情報を示す．

このうち属性値に欠損のある観測データは 219 件であり，まずこれを 4.1.1 節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 分割交差検証を行った結果は，正解率 69.7，再現率 70.9%，F 値 69.7%であった．表 4.2 に Confusion Matrix を示す．結果として，観測データ数の少ない分類 3(18〜35 台) や分類 4(36〜100 台) の推定精度が低いため，全体としての精度が低くなったが，その他の分類の推定精度はある程度高い結果が得られた．

表 4.1: 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	17	673
東京都練馬区	3	96
神奈川県横浜市中区	1	39
東京都福生市	2	23
東京都府中市	3	20
東京都武蔵野市	3	16
北海道札幌市中央区	3	14
東京都新宿区	1	4
神奈川県横浜市磯子区	2	3
東京都国分寺市	2	3
東京都渋谷区	1	3
北海道札幌市北区	1	2
東京都品川区	1	1

そして，897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．この推定台数を，図3.6のように独自に定義した`ipb:estimatedValue`プロパティを用いてLODに追加することで，放置自転車LODの時間的な欠損を補完する．

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．本研究では，各観測地点において初観測日から最終観測日までの間で，午前9時または午後9時に観測がされていない場合に，前述のパラメータ設定で放置自転車の台数を推定する．推定値とその確率を基に次のようなRDFを生成し，放置自転車LODに追加する．

```
@prefix ipb: <http://www.ohsuga.is.uec.ac.jp/ipblod/vocabulary#>
@prefix bicycle: <http://www.ohsuga.is.uec.ac.jp/bicycle/resource/>
bicycle:ipb_{observation point}_{datetime}
    ipb:estimatedValue [
        rdf:value "0-6" ;
        ipb:probability "0.772"^^xsd:double
    ] .
```

表 4.2: Confusion Matrix

		分類 1	分類 2	分類 3	分類 4
推定値	分類 1	339	38	10	2
	分類 2	62	158	19	2
	分類 3	22	34	89	2
	分類 4	2	16	5	5
TP		0.871	0.656	0.605	0.179

4.1.5 考察

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．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．そのため，観測地点を過度に細分化して報告されていたデータがあり，ベイジアンネットワークの精度に影響を与えたと考えられる．さらに，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．

本研究ではドメイン要件と実用性に焦点を当ててナレッジグラフのスキーマ設計の方法論と事例としての放置自転車スキーマを示した．ドメイン要件に関しては，今回，放置自転車をイベントとして捉えることにしたため，放置自転車問題を EO に適切に対応させる

ことがドメイン要件充足の一つの指標として見ることができる．提案手法では，EO のプロパティに基いてキーワードを抽出し，これをクラスタリングした結果を基にスキーマを設計し，実データを取得できない agent と product などを除き，EO に沿ったナレッジグラフを構築した．更にその際，プロパティの domain（定義域）と range（値域）の制約にも違反しないことを考慮しており，EO との意味的整合性も保持している．したがって，放置自転車をイベントとして捉えた時のドメイン要件を，現実的に可能な範囲で満たしておりスキーマとして適切であると考える．

また，EO に基づいて設計したスキーマを使用して属性選択を行い，ベイジアンネットワークにより約 70%の精度で時間的欠損値を推定できたことは，放置自転車の要因に関してスキーマの網羅性はある程度充分であり，放置自転車のスキーマ設計に EO を採用したことは妥当であったと考えられる．ベイジアンネットワークに使用した属性は，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，降水の有無，駐輪場料金の有無，周辺施設のタイプである．これは属性値を実験的に増減させたときに時間的欠損値推定精度が最も良くなった例である．本研究では放置自転車に関して，Google から記事検索を行った上位 10 件の記事とその参考文献を要因の抽出範囲とし，その中でもデータとして追加することが可能なものを対象とした．今後，道路の工事日程，ニュータウン開発や駅のダイヤ変更などの新たな情報を追加することができれば，精度向上に貢献する可能性がある．

本研究では，時間的な欠損値の推定について放置自転車問題を対象としたが，提案手法は都市の地理空間的な特徴と人の行動に関連の深い同様な都市問題という対象範囲内で部分的に一般化できる．ベイジアンネットワークで使用する属性はナレッジグラフのスキーマに依存し，対象とする範囲の都市問題の性質上，スキーマに地理情報や日時情報など放置自転車問題と共通する情報が含まれることが考えられる．その他，個々の都市問題により異なる原因がスキーマに組み込まれることが考えられ，使用する属性によって推定精度に差が出るため，この点は都市問題ごとに調整が必要である．したがって，スキーマ設計の方法論は同様の都市問題に対して一般化でき，個々の都市問題ごとに構築されるナレッジグラフの構造は大きくは変わらない．また，都市問題ごとにスキーマに含む情報に多少独自要素が入ることが予想され，ベイジアンネットワークで使用する属性値にも変更が生じるが，同様の問題かつ時系列データが用意できれば時間的欠損値推定の一連の流れは一

般化できる．

4.2 流体シミュレーションを用いた空間的な欠損値の推定

前節で述べたように，放置自転車 LOD はソーシャルセンサから収集したデータを基に構築しているため，時空間的な欠損値が生じている．時間的な欠損値に関してはベイジアンネットワークを用いて 69.7 の精度で補完しているが，空間的な欠損値（放置自転車があるが未観測の地点）については課題となっていた．そこで，本節ではこの空間的な欠損値を補完することにより，放置自転車 LOD を地理空間的に拡充する手法を説明する．

これまでの観測データから，放置自転車の存在する地点には建物の密集具合や影，道路の幅などの地理空間的な共通点が存在すると考えた．そこで，都市に流体を流し込んだときに生じるよどみ点と，放置自転車の観測地点にある程度相関があるとの仮説を立て，流体シミュレーションを用いた空間的欠損値の推定手法を提案した．流体シミュレーションでは建物が密集している地点や高い建物の影などによどみが発生しやすく，建物が少ない地点ではよどみが発生しにくい考えられる．したがって，都市空間の複雑な構造と人の流れとの関係を擬似的な数値情報として得ることが可能であるため，本研究では流体シミュレーションを使用するという着想に至った．

人の避難行動のシミュレーション研究においては物理モデル手法，セルオートマトン手法，エージェントシミュレーションが存在する [74]．セルオートマトン法とエージェントシミュレーションは行動ルールを設定する必要があり，放置自転車のように個人の目的が不明瞭な都市問題においては，これらの手法を直接取り入れることは難しい．そこで，本研究では流体の保存則に従う物理モデルを用いて放置自転車のシミュレーションを行う．具体的には，駅付近の 3D 地図を基に気流のシミュレーション行うことでよどみ点を検出し，放置自転車の観測地点との相関性を調べた．さらに DBpedia Japanese [32] を用いてよどみ点をフィルタリングし，この結果を放置自転車が存在する可能性の高い地点として，空間的な欠損値を推定した．

4.2.1　流体シミュレーションを用いたよどみ点の検出

気流のシミュレーション手法には風洞実験と数値流体力学 (Computation Fluid Dynamics, CFD) がある．CFD とは，流体に関する方程式をコンピュータで計算することで流体の動きを観測する手法である．大型で高価な実験装置が必要である風洞実験に対し，CFD はコンピュータを使用するため環境を変えながら実験することが容易である．しかしながら，近似解を用いるため流体の動きを完全には再現できないという欠点がある．

本研究では国土地理院基盤地図情報サービス[2]から建物の平面データを取得し，地理情報サービスの ArcGIS[3]を用いて建物の 3D データを作成した．図 4.1 は調布駅付近の建物データの上に，放置自転車の観測地点をマッピングした図である．赤いマーカは放置自転車の観測地点であり，放置自転車 LOD の SPARQL エンドポイントから取得したものである．次に，ArcGIS 上で動作する気流のシミュレーションソフトである Airflow Analyst[4]を使用して，駅付近の気流シミュレーションを行った．図 4.2 は調布駅を中心として解析範囲を設定して格子を生成した図である．解析範囲は放置自転車の観測地点が全て含まれるように設定した．図 4.2 では調布駅を中心として 700m 四方を解析範囲とし，格子点間の距離を 5m，合計格子点数を 10,000 としている．また，風源は駅に伸びる道路と平行になるように設定し，調布駅の場合は道路に沿って 4 方向から人が駅に向かうことを想定し 11 度，109 度，190 度，288 度と風向を変えてシミュレーションを行う．

図 4.3 は風向を 11 度として，風速 5m/s の風を発生させた時のシミュレーション結果を基に，全格子点における平均風速の値を OpenStreetMap[5]上にプロットした結果である．青い点が格子点であり，この点の大小が平均風速を表している．この数値データを基に気流のよどみ点を検出する．よどみ点とは流れ場の中で流体の速度が 0 になる点である．我々は図 4.4 のパターンを全格子点に適用してよどみ点の検出を試みた．黒色部分は平均風速が 0 以上の点，白色部分が平均風速が 0 の点，灰色部分が平均風速が 0 より大きく 0.1 以下の点である．実際には平均風速が 0 の格子点は建物がある点となるため，本研究では隣接する格子点より平均風速が低く，かつ平均風速が 0.1 以下の点をよどみ点として定義し

[2]http://www.gsi.go.jp/kiban/
[3]http://www.esrij.com/products/arcgis/
[4]http://www.airflowanalyst.com/
[5]https://openstreetmap.jp/

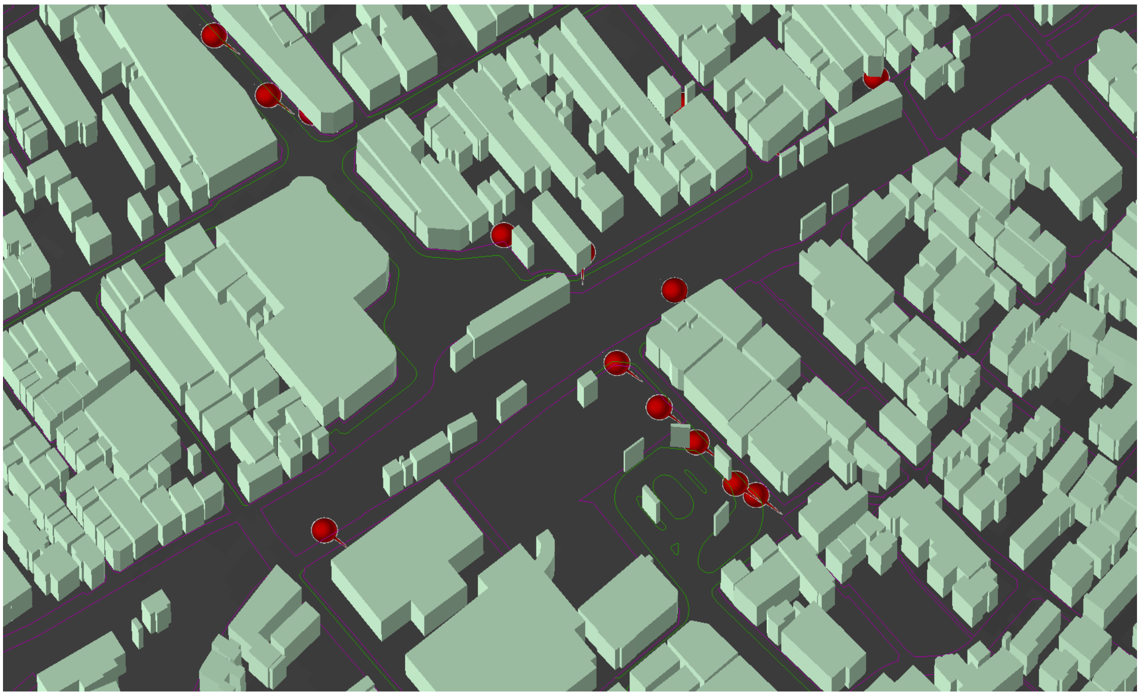

図 4.1: 調布駅周辺の 3D 地図

た．全よどみ点の内，放置自転車の観測地点から半径 20m 以内に含まれるよどみ点の割合を適合率と定義した．また，全観測地点の内，半径 20m 以内によどみ点を含む観測地点の割合を再現率とした．調布駅，府中駅，新宿駅において全てのよどみ点パターンで実験を行った結果を表 4.3 に示す．結果として，パターン (a) と (j) が最も放置自転車の観測地点の分布と近い結果となったため，本研究においてはよどみ点の検出にパターン (a) と (j) を使用する．図 4.5 は調布駅において 4 方向のシミュレーション結果に図 4.4 のパターン (a) と (j) でよどみ点を検出し，4 方向の検出結果をマージして地図上にプロットしたものである．青いマーカはよどみ点であり，赤いマーカは放置自転車の観測地点である．

4.2.2 セマンティクスを用いたよどみ点のフィルタリング

前節で述べた手法によりよどみ点を検出したが，図 4.5 を見て分かる通り多くのノイズが含まれている．我々は，気流のよどみ点付近であり，かつ注目度の高い POI が周囲に存在する点に放置自転車が発生しやすいと考えた．そこで，よどみ点周囲の POI の注目度を DBpedia Japanese を用いて算出し，注目度の総和でフィルタリングした．

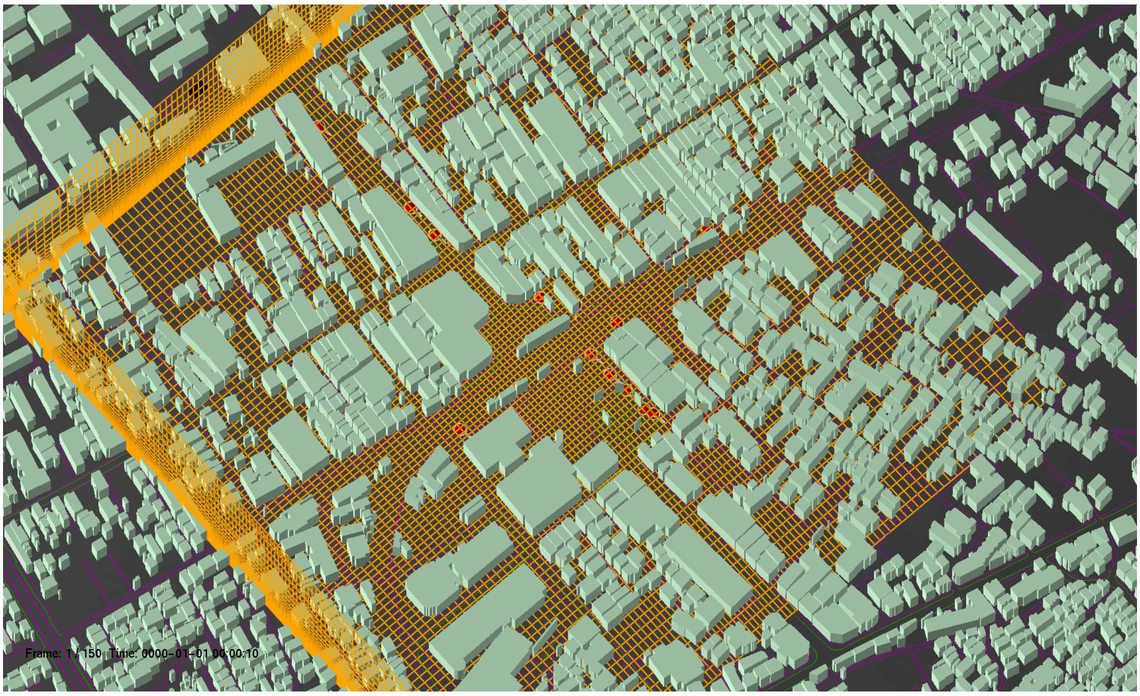

図 4.2: 調布駅周辺の格子生成図

まず，前節で得られたよどみ点の緯度経度を基に，Google Places API で半径 20m 以内の POI を検索する．取得した POI の type に関するリソースについて，DBpedia Japanse 上での被リンク数を計算する．Google Places API のタイプは DBpedia Japanse のリソースに予めマッピングしている．我々は，人物に関するリソースからの被リンク数を注目度とし，POI の type に関するリソースが `foaf:Person` のインスタンスから受けているリンクの数を取得した．その後，周辺 POI の注目度の総和を計算し，注目度の合計が 200 未満であるよどみ点を除外する．この閾値は 100 から 1000 まで変動させて，最も結果が良くなったものを使用している．図 4.6 に図 4.5 のフィルタリング後の結果，すなわち最終的な空間的欠損の推定値を示す．

さらに，この空間的欠損の推定値を以下の RDF 表現で放置自転車 LOD に追加した．

```
@prefix ipb: <http://www.ohsuga.is.uec.ac.jp/ipblod/vocabulary#>
@prefix bicycle: <http://www.ohsuga.is.uec.ac.jp/bicycle/resource/>
@prefix geo:   <http://www.w3.org/2003/01/geo/wgs84_pos#> .
@prefix ogcgs: <http://www.opengis.net/ont/geosparql#> .
@prefix ngeo:  <http://geovocab.org/geometry#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
```

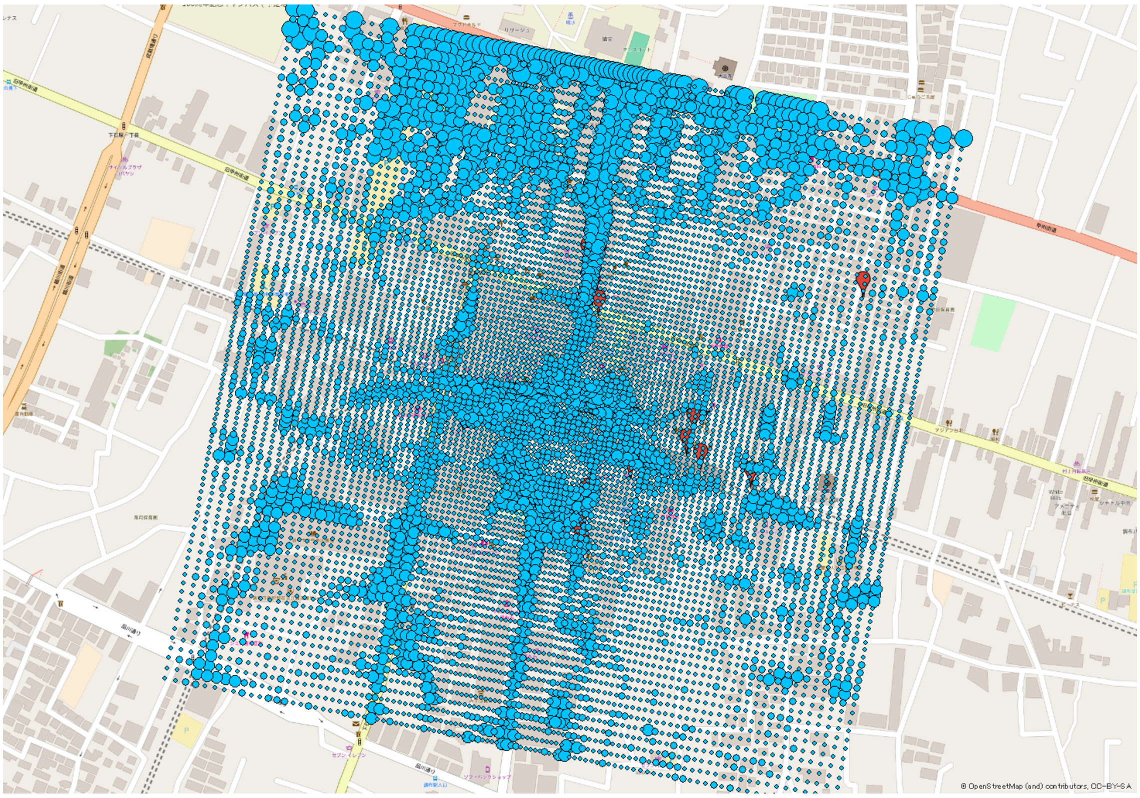

図 4.3: 調布駅周辺の平均風速

```
@prefix gn:    <http://www.geonames.org/ontology#> .
@prefix gnjp:    <http://geonames.jp/resource/> .
@prefix xsd:   <http://www.w3.org/2001/XMLSchema#> .
bicycle:estimated_obs_{timestamp}
    rdf:type     ipb:EstimatedObservationPoint ;
    geo:lat      "latitude"^^xsd:double ;
    geo:long     "longitude"^^xsd:double ;
    gn:parentADM          gnjp:{Prefecture}
    gn:parentADM2         gnjp:{City, Prefecture} ;
    gn:parentADM3         gnjp:{Town, City, Prefecture} ;
    gn:parentADM4         gnjp:{Land lot, Town, City, Prefecture} ;
    ngeo:geometry         [ a ngeo:Geometry; ogcgs:asWKT  "POINT(
        latitude,longitude)"^^<http://www.openlinksw.com/schemas/
        virtrdf#Geometry> . ] ;
    gn:nearby    bicycle:{POI name} ;
    dcterms:created      "datetime"^^xsd:dateTime .
```

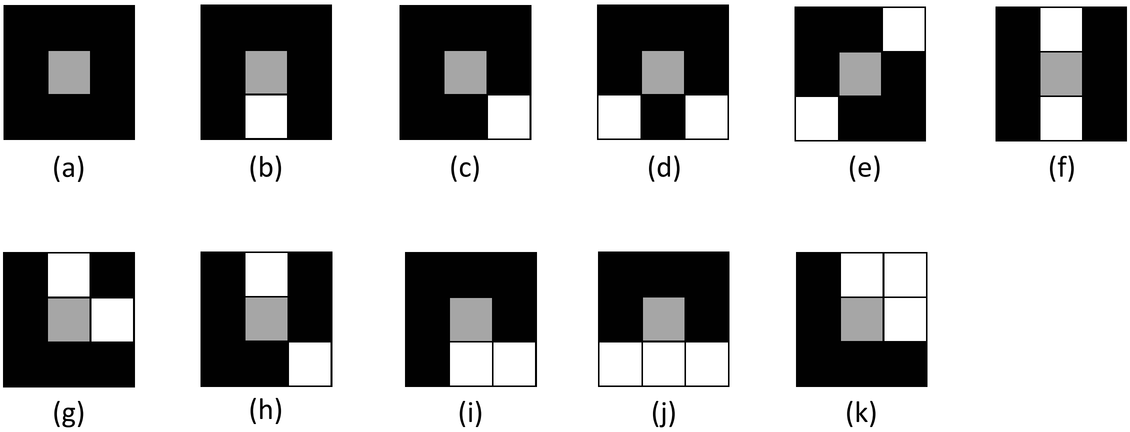

図 4.4: 本研究で定義したよどみ点のパターン

表 4.3: 各よどみ点パターンの精度

パターン	適合率	再現率	F 値
(a)	0.102	0.286	0.150
(b), (c)	0.0833	0.0357	0.0500
(d), (e), (f), (g), (h), (i)	0.000	0.000	0.000
(j)	0.0913	0.429	0.151
(k)	0.0746	0.107	0.0880

4.2.3　評価

本節では，4.2.1 節で述べたよどみ点検出手法と 4.2.2 節で述べた手法それぞれとベースライン手法を比較することで，本研究の提案手法の優位性について検証した結果を示す．

本研究では，事前に放置自転車の観測地点を複数取得できている調布駅，府中駅，新宿駅を対象に実験を行った．放置自転車の観測地点数は 56 地点であった．まず，4.2.1 節で述べたよどみ点検出手法（以下，提案手法 1）とベースライン手法を比較する．ベースライン手法では一様性を保つために，提案手法 1 により推定された地点数とほぼ同数となるように格子の幅と格子点数を設定し，この格子点をベースライン手法における推定地点としている．図 4.7 に本研究のベースライン手法を示す．表 4.4 にベースライン手法と提案手法

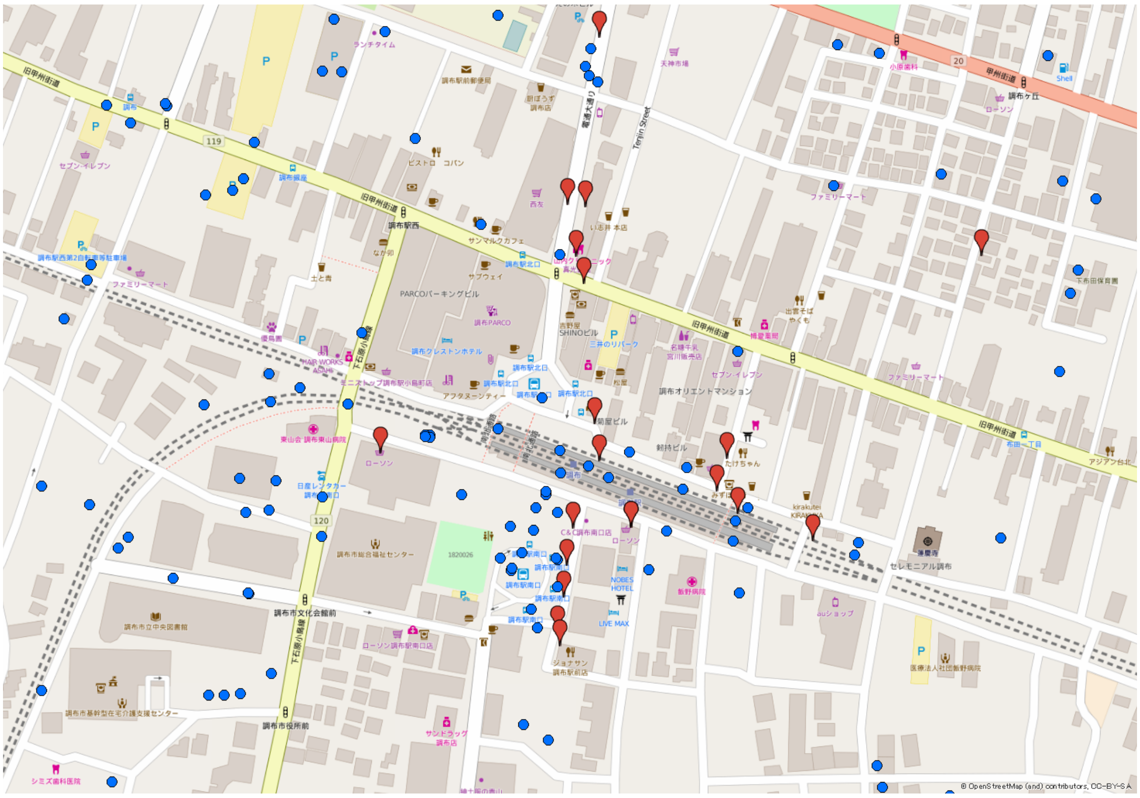

図 4.5: 調布駅周辺にけるよどみ点の検出結果

1 の再現率，適合率，F 値を示す．結果として，提案手法 1 は再現率，適合率，F 値の全てにおいてベースライン手法を上回った．また，ベースライン手法と提案手法 1 の間で再現率および適合率に差がないとの帰無仮説のもと，正解値と不正解値からなる 2x2 行列に対して有意水準 0.05 で χ 二乗検定を適用した．その結果，p 値は 0.0159 となり帰無仮説は棄却された．従ってベースライン手法と提案手法 1 に優位な差が見られた．すなわち，CFD の結果から提案したパターン (a) と (j) を用いてよどみ点を検出する手法は，放置自転車の新規観測地点検出の手法として一定の優位性があると考える．

次に，4.2.2 節で述べた，提案手法 1 と DBpedia Japanese を用いたフィルタリングを組み合わせた手法（以下，提案手法 2）とベースライン手法を比較する．提案手法 2 は提案手法 1 より推定地点数が減っているため，今回使用するベースライン手法では，推定地点が提案手法 2 とほぼ同数になるように格子の幅を調整している．そのため，表 4.4 のベースライン手法とは結果が異なっている．表 4.5 にベースライン手法と提案手法 2 の再現率，

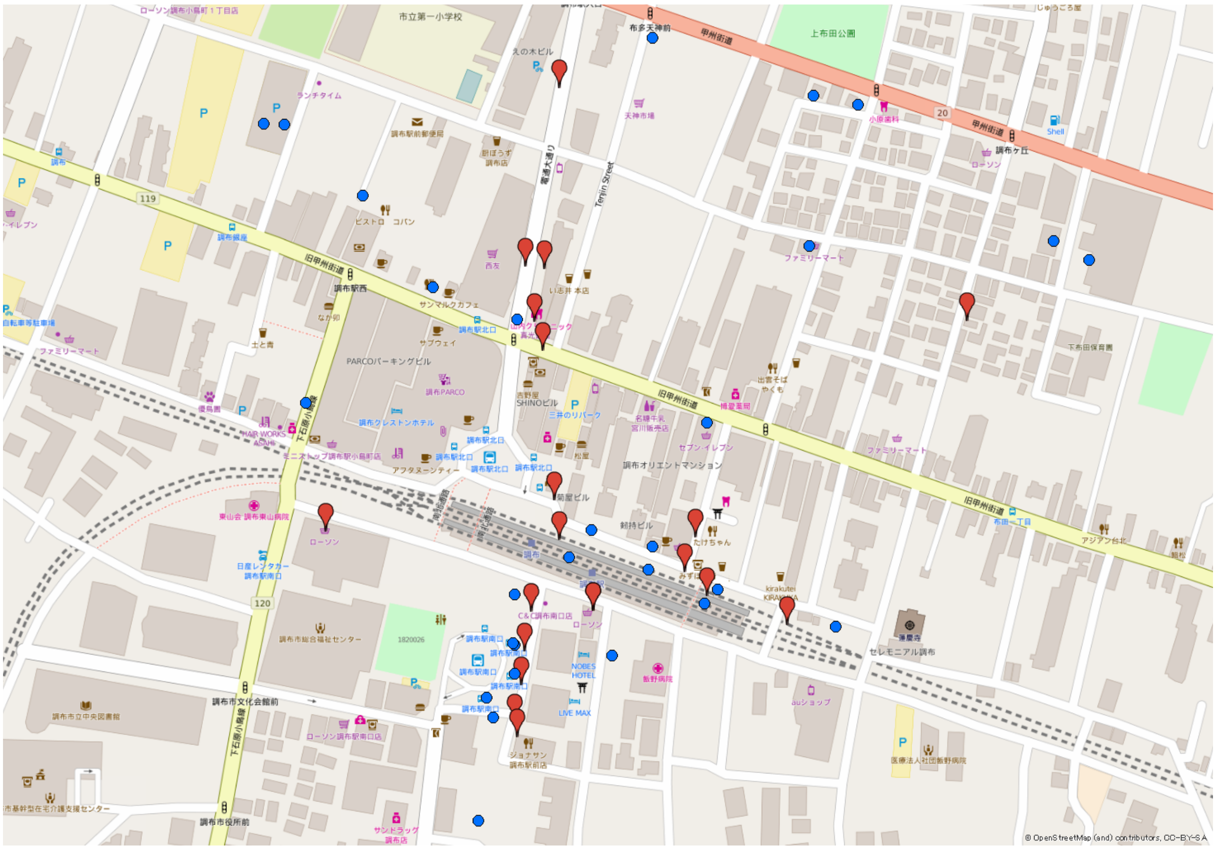

図 4.6: 調布駅周辺におけるフィルタリング後のよどみ点（最終的な推定値）

適合率，F 値を示す．提案手法 1 と同様に，提案手法 2 も再現率，適合率，F 値の全てにおいてベースライン手法を上回る結果となった．また，表 4.4 と表 4.5 から，提案手法 2 は提案手法 1 より適合率と F 値が高くなることがわかった．さらに，提案手法 2 とベースライン手法の間で再現率および適合率に差がないとの帰無仮説のもと，有意水準 0.05 で χ 二乗検定を適用した．その結果，p 値が 0.0000739 となり帰無仮説は棄却された．従ってベースライン手法と提案手法 2 に優位な差が見られ，DBpedia Japanese を用いてよどみ点をフィルタリングし，空間的欠損の推定値とする手法は一定の有用性があると考える．

さらに，相互最近隣距離法 (Mutual Nearest Neighbor, MNN) を用いて推定地点と観測地点の点分布が近接傾向にあるか検証する．相互最近隣距離法は空間解析の分野で用いられる手法であり，異なる点分布間の最近距離を計算し，ランダムに分布した場合の結果と比較することで，二つの点分布が近接傾向にあるかランダムであるか，回避傾向にあるかがわかる．MNN における平均距離は次式で求める．

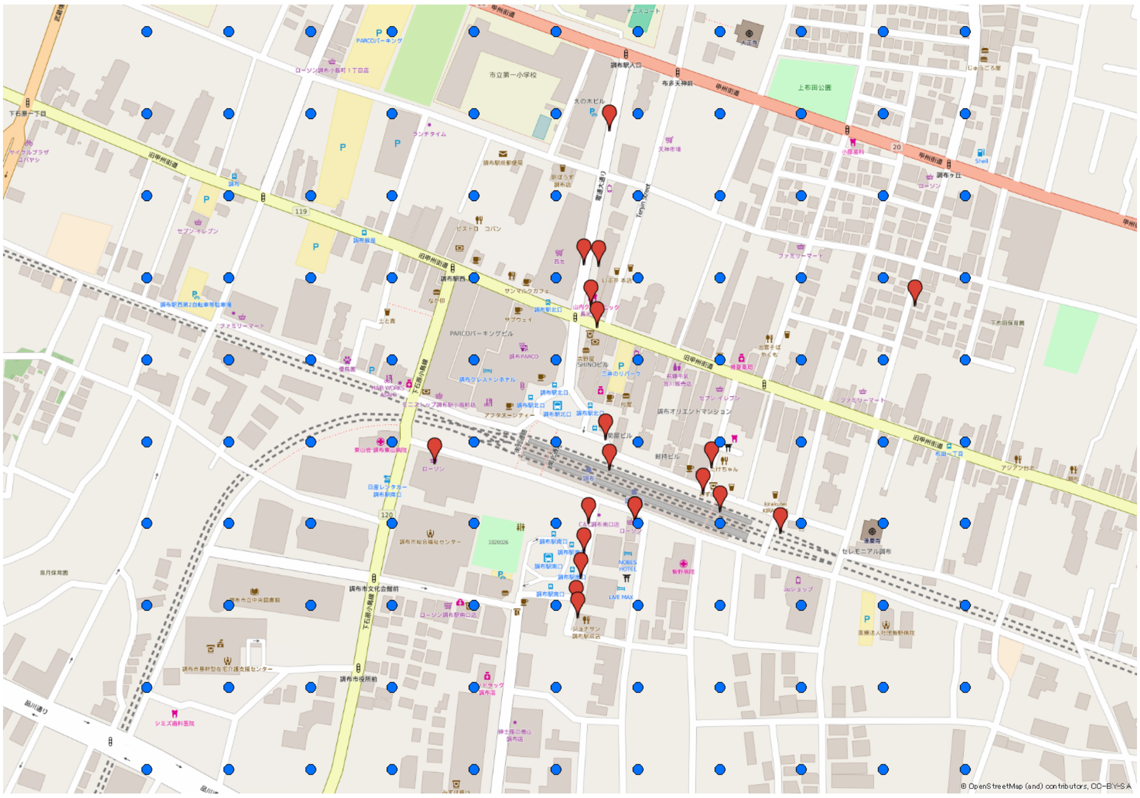

図 4.7: ベースライン手法

$$D = \frac{1}{n_a + n_b}\left(\sum_{i=1}^{n_a} d_{ai} + \sum_{i=1}^{n_b} d_{bi}\right) \tag{4.2}$$

点分布 A と点分布 B があるとき，分布 A (または B) 内の点から最も近い分布 B (または A) 内の点までの距離は d_{ai} (または d_{bi}) である．n_a (または n_b) は点分布 A (または B) 内の点の数である．点分布 A と B がランダム分布のとき，MNN の平均距離は次式で予想される．

$$E[D] = \frac{1}{n_a + n_b}\left(\frac{n_a}{2\sqrt{\lambda_b}} + \frac{n_b}{2\sqrt{\lambda_a}}\right) \tag{4.3}$$

したがって，本検証においては図 4.7 のベースライン手法の代わりにランダム分布を採用した．ここで，λ_a (または λ_b) は A (または B) の密度である．結果として，$E[D]$ は 61.4 となり，提案手法の D の値は 46.9 となった．自由度は 88，有意水準は 0.01 である．t 検定

表 4.4: ベースライン手法と提案手法 1 の推定精度

		ベースライン手法	提案手法 1
調布駅	適合率	0.049	0.124
	再現率	0.231	0.539
	適合率	0.081	0.202
府中駅	適合率	0.125	0.231
	再現率	0.222	0.444
	F 値	0.160	0.304
新宿駅	適合率	0.049	0.167
	再現率	0.190	0.667
	F 値	0.078	0.267
Total	適合率	0.0550	0.150
	再現率	0.214	0.571
	F 値	0.087	0.238

の結果，t 境界値は 2.633，t 値は 4.76 であった．したがって，提案手法の結果とランダム推定手法の結果には有意な差があることがわかった．また，提案手法により推定された点分布と観測データの点分布は近接傾向にあることがわかり，観測地点と推定地点は相関があるといえる．したがって，流体のよどみ点と放置自転車の観測地点にはある程度相関があるとの仮説は正しいことが確認できた．

4.2.4　考察

提案手法の精度が落ちた原因として，観測地点数の不足が考えられる．本研究で正解データとして使用した観測値点数は調布駅，府中駅，新宿駅付近の合計で 56 地点である．実際には提案手法により推定された地点の付近にも放置自転車が存在している可能性はあり，正確な精度を計算するためには推定地点の実地調査による確認が必要である．また，本手法では公有地や私有地などの個人が立ち入ることができない空間も推定値の対象となっていたため，これらの範囲内で得られた推定値がノイズとなっていることがわかった．今後は公有地，私有地などのデータを用いてさらに推定値をフィルタリングする必要があると考えている．

表 4.5: ベースライン手法と提案手法 2 の推定精度

		ベースライン手法	提案手法 2
調布駅	適合率	0.0469	0.247
	再現率	0.115	0.462
	F 値	0.0667	0.322
府中駅	適合率	0.125	0.250
	再現率	0.222	0.333
	F 値	0.160	0.286
新宿駅	適合率	0.0493	0.211
	再現率	0.190	0.571
	F 値	0.0784	0.308
Total	適合率	0.0559	0.228
	再現率	0.161	0.482
	F 値	0.0829	0.310

本手法の適用対象として放置自転車問題を例とした．より一般化すると，都市の地理空間的な特徴と人の行動に関連の深い都市問題のうち，特に発生地点が空間的にまばらである問題において適用が可能である．同様の問題として，例えば空き巣や痴漢などの問題に本手法が適用できると考える．ただし，気流シミュレーションにおけるパラメータやよどみ点のパターンや，フィルタリングの際に使用する注目度などは，個々の問題によって調整が必要となる．

4.3 Graph Embedding とニューラルネットワークによる空間的欠損値の推定

ポイ捨て調査結果の LOD については，調査地域におけるある一時点でのゴミのポイ捨て分布を表している．したがって時系列データとしては乏しく，放置自転車の時のようにベイジアンネットワークを用いた時間的な欠損補完を行うことや，その他機械学習のアプローチにより時系列データ予測の手法を適用することは困難である．一方で，地理空間的データとしては各調査地域内では密なデータであるものの，一度の調査に多くの時間と労

力が必要であるために調査地域数は多くない．そこで，ポイ捨て調査結果の LOD の拡充については，空間的欠損値（未調査地域）のデータを推定することを目的とする．

本節では，建物，道路，POI などの都市の地理空間的な特徴と，人を流体と見立てた流体シミュレーション結果をもとに，ニューラルネットワークを用いてゴミのポイ捨て分布を推定する．本研究では都市をグリッド分割し，各セル内におけるポイ捨てゴミの個数を次の 3 つの手法を用いて推定して結果を比較する．

(1) 画像ベース手法：地理空間的な特徴とシミュレーション結果の地図画像をもとに畳み込みニューラルネットワーク (CNN) でゴミの数を推定

(2) グラフベース手法：地理空間的な特徴とシミュレーション結果をナレッジグラフ化し，Graph Embedding を行ってからニューラルネットワークによりゴミの数を推定

(3) 画像-グラフ結合手法：(1) と (2) のニューラルネットワークモデルを結合してゴミの数を推定

画像ベース手法は直感的な手法であり，画像にすべての情報を詰め込み，一般的に画像の分類や回帰予測のタスクで高精度な手法である CNN を使用する．グラフベース手法は，情報を一度知識構造化してからベクトル空間に埋め込むことで，高い精度で推定できるという仮説に基づいている．画像ベース手法とグラフベース手法を図 4.8 に示す．画像-グラフ結合手法では，画像ベース手法とグラフベース手法を組み合わせることで特徴を相補的に学習し，高い精度での推定が期待できると考えた．

都市における予測問題としては人口移動 (crowd flow) の予測がよく知られたタスクである．ST-ResNet [71] では都市をグリッド分割して人口流入と人口流出をテンソルで表現し，複数の Residual Unit から構成される Deep Residual Neural Network [24] により市の人口移動状況を予測する．FCCF [26] では道路ネットワークを使用して都市を地域ごとに分割し，Intrinsic Gaussian Markov Random Field (IGMRF) [52] とベイジアンネットワークを使用して人口移動を予測している．また，都市の犯罪データをもとに Deep CNN を用いて新規時刻における犯罪を予測する研究がある [17]．これらの研究は数年単位で蓄積された時間的かつ空間的データを対象としており，ある時刻における都市の状況を予測するものである．しかしながら，我々の知る限り都市単位でゴミのポイ捨てを継続的に調査して蓄

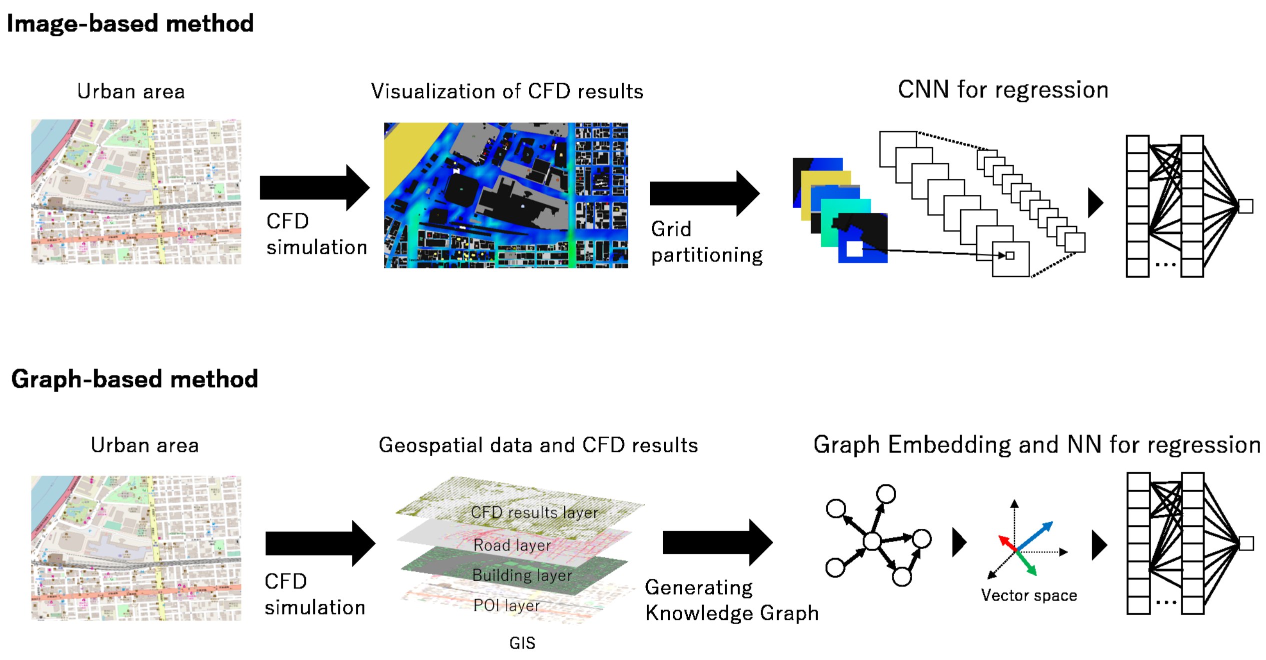

図 4.8: 空間的欠損値の画像ベース推定手法とグラフベース推定手法の流れ

積されたデータは存在しない．また，時系列情報の乏しい空間データのみを使用した都市問題の予測手法およびナレッジグラフの空間的欠損値の推定手法も存在しない．そこで，本研究では空間データのみを使用して都市問題を予測し，ナレッジグラフの空間的な欠損値を推定する手法を提案する．これにより，少ない調査データで他地域への横展開し，ナレッジグラフの空間的な拡充が可能になる．

4.3.1 流体シミュレーション結果と地理空間情報の Graph Embedding による推定

本研究では GIS ソフトの ArcGIS 上に，国土地理院から取得した地理標高データと，ArcGIS Geo Suite 詳細地図[6]の建物データ，道路縁データを表示し，4.2.1 節と同様に Airflow Analyst を使用して気流シミュレーションを行った．図 4.9 は両国駅を中心として格子を生成した図である．CFD の結果が収束するように格子の数，間隔，範囲を設定する．また，4.2.1 節と同様に，風源は駅に伸びる道路と平行になるように設定する．図 4.10 は風向を

[6]https://www.esrij.com/products/data-content-geosuite-shosai/

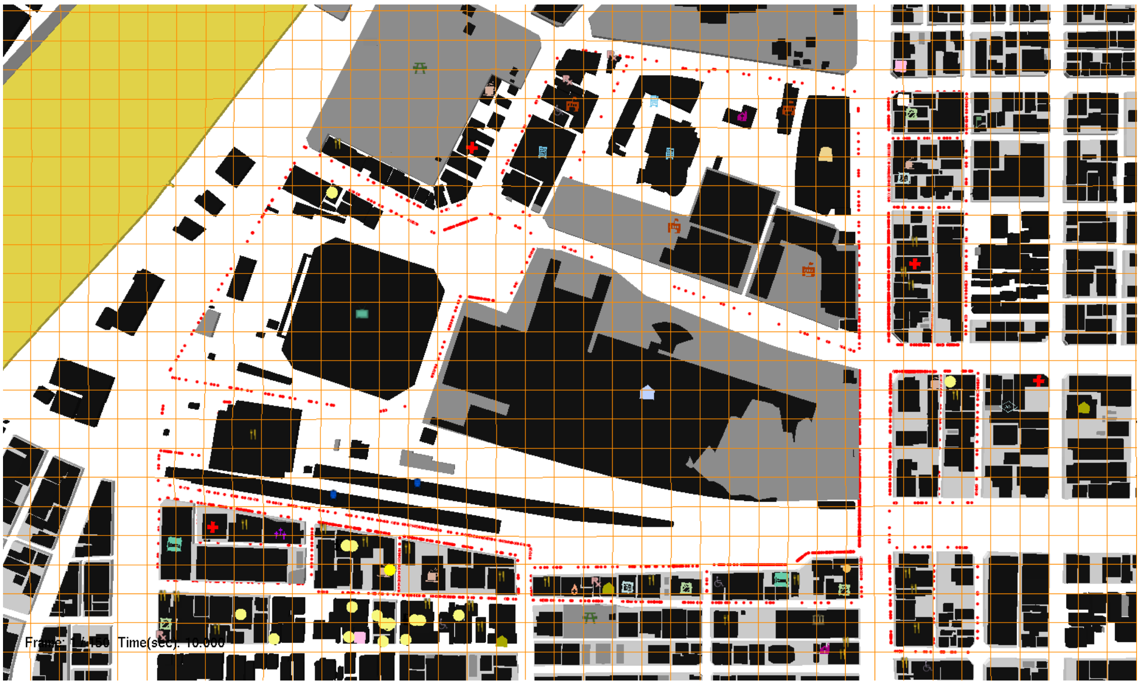

図 4.9: 両国駅を中心とした格子生成図

15 度として，風速 5m/s の風を発生させた時のシミュレーション結果をもとに，格子点における平均風速の値を地図上にプロットした結果である．青い点が格子点であり，この点の大小が平均風速を表している．この数値データをもとにナレッジグラフを構築する．

本手法のナレッジグラフ化のイメージを図 4.11 に示す．格子線で囲まれたセルごとにリソースを作成し，各セルが 4 つの格子点を持つ．各格子点は CFD の結果をリテラル情報として持つ．本研究では CFD の結果の内，平均風速と平均風向のみを使用する．また，セル内に建物の一部分が含まれる場合は建物の部分リソースを作成し，`have` プロパティでリンクする．建物の部分リソースは `partOf` プロパティで建物全体のリソースとリンクする．同様に道路に関してもリソースを作成し，セルと関連付ける．建物が POI である場合はその種類を持つ．

本手法の目的は，格子分割された都市において，各セルに含まれるポイ捨てゴミの個数を予測することである．そこで，各セルのベクトル表現を Graph Embedding 手法を用いてナレッジグラフから生成する．我々は Graph Embedding 手法として RDF2vec，node2vec，TransE，TransR を比較し，最も精度の良くなった RDF2vec を使用した．各手法の比較に

図 4.10: 両国駅の格子点における平均風速のプロット図

ついては 4.3.4 節で後述する．RDF2vec は RDF からシーケンス集合を生成し，このシーケンス集合をもとに word2vec によりベクトル表現を獲得する．まず，グラフ $G = (V, E)$ が与えられたとき，各頂点 $v \in V$ に関してルート v_r を起点に深さ d のグラフウォークパス P_v を $v_r \to e_{1,i} \to v_{1,i} \to \cdots \to v_{d,i}$ のように生成する．ここで $i \in E(v_r)$．次に，Weisfeiler-Lehman Subtree RDF Graph Kernels (WLRDF) のグラフラベリング手法を用いて，頂点とエッジに数字ラベルを割り当てる．WLRDF の各反復において，リラベルされたグラフ上でウォーク P_v をたどることで，各要素がリラベルされたウォークパスを生成する．各ウォークパスの先頭には v_r を追加する．各頂点におけるすべてのウォークを WLRDF の最大反復回数 h まで生成し，最終的なシーケンス集合は $\bigcup_{i=1}^{h} \bigcup_{v \in V} P_v$ となる．

word2vec では Continuous Bag-of-Words (CBOW) モデルと Skip-Gram モデルの 2 つのアルゴリズムが使用される．CBOW モデルは文脈の window 幅 c で出現する単語 $w_{t+j}(-c \leq j \leq c)$ が与えられたとき，単語 w_t が得られる平均対数確率 (式 4.4) を最大化する．

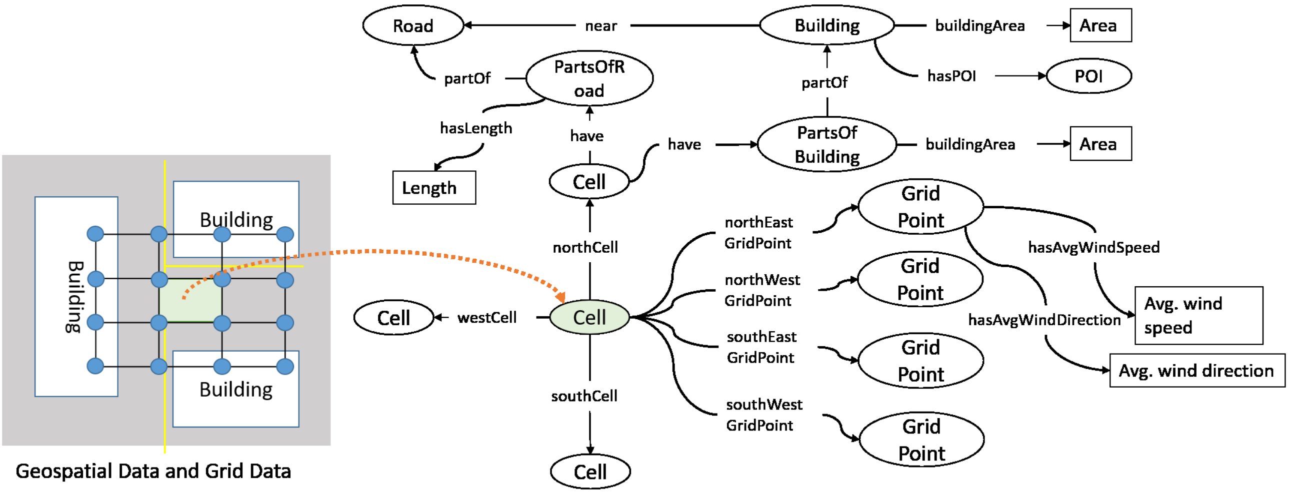

図 4.11: 流体シミュレーション結果と地理空間情報のナレッジグラフ化のイメージ

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{j} log(p(w_t|w_{t+j})) \tag{4.4}$$

ここで，確率 $p(w_t|w_{t+j})$ は softmax 関数を用いて次式で計算される．

$$p(w_o|w_i) = \frac{exp(\overline{v}^T v'_{w_o})}{\sum_{w=1}^{V} exp(\overline{v}^T v'_w)} \tag{4.5}$$

ここで，w_i と w_o はそれぞれ入力単語と出力単語，v_w はそれらの入力ベクトル表現，v'_w は出力ベクトル表現，V は単語集合である．また，$\overline{v}$ はすべての文脈単語の入力ベクトルの平均として次式で計算される．

$$\overline{v} = \frac{1}{2c}\sum_{-c\leq j\leq c, j\neq 0} v_{w_{t+j}} \tag{4.6}$$

Skip-Gram モデルは CBOW を逆向きにしたモデルである．単語 w_t が与えられたとき，文脈の window 幅 c で出現する単語 $w_{t+j}(-c \leq j \leq c)$ が得られる平均対数確率 (式 4.7) を最大化する．

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{j} log(p(w_{t+j}|w_t)) \tag{4.7}$$

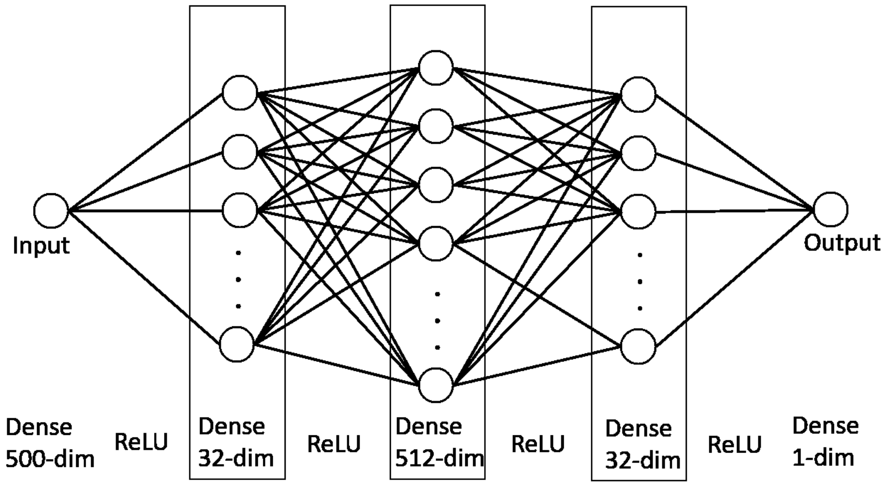

図 4.12: Graph Embedding 後にゴミの個数の予測に使用する全結合ニューラルネットワークのモデル

ここで，確率 $p(w_{t+j}|w_t)$ は softmax 関数を用いて次式で計算される．

$$p(w_o|w_i) = \frac{exp(v_{w_o}^{'T} v_{w_i})}{\sum_{w=1}^{V} exp(v_w^T v_{w_i})} \tag{4.8}$$

RDF2vec においては，Skip-Gram モデルのほうが精度が高くなることが分かっているためこちらを使用する [51]．前述のグラフウォークステップで生成されるシーケンス集合を擬似的な文章として入力することで，RDF における各頂点とエッジごとにベクトル表現を得ることができる．

さらに，Cell リソースのベクトル表現とセル内のポイ捨てゴミの個数の正解ラベルをもとに機械学習し，未知のセルにおけるポイ捨てゴミの個数を推定する．本研究では機械学習手法として全結合ニューラルネットワークを使用し，図 4.12 のモデルで予測する．

RDF2vec ではリテラル情報を扱うことができないため，本研究ではリテラル情報を等間隔分類して擬似的なリソースを作成する．例えば平均風速は最小値 0 から最大値 5.5 までの値で 12 分割し，インスタンスを作成している．平均風向については 8 方向に分類している．その他，建物面積や道路の長さも等間隔分類している．

さらに，RDF2vec では各プロパティの逆プロパティが存在する場合に，グラフウォークにおいて局所的な歩行で終了する場合がある．例えば，$A \xrightarrow{northCell} B$ と $B \xrightarrow{southCell} A$ は

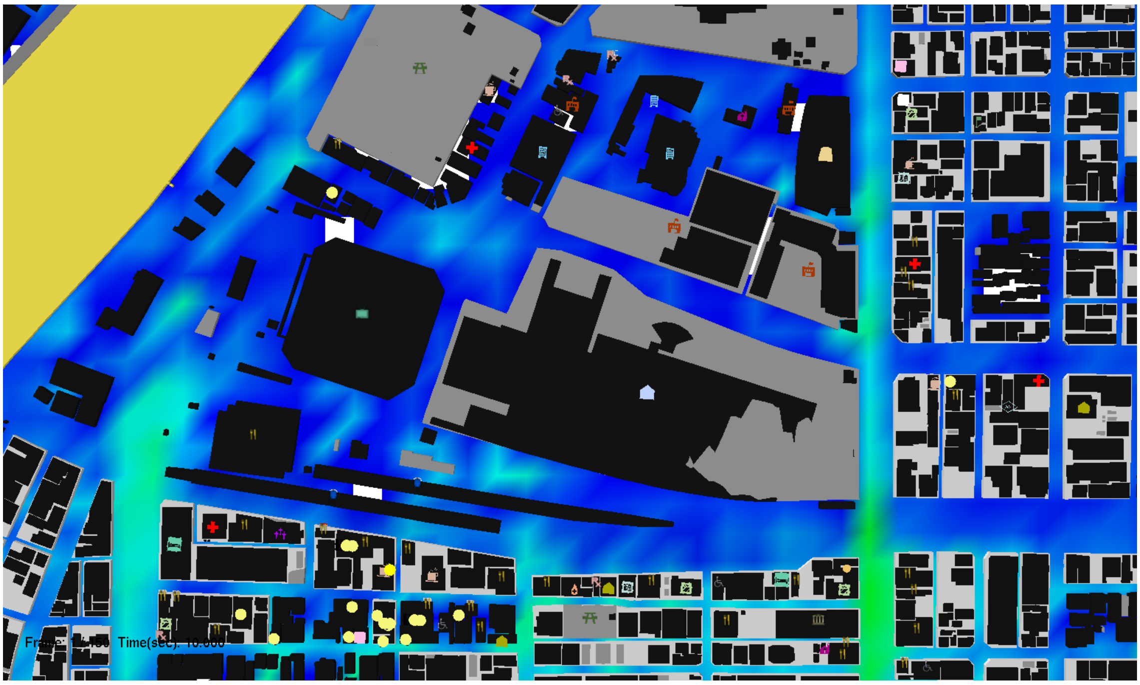

図 4.13: 両国駅の CFD シミュレーション結果のヒートマップ表示画像

逆関係であり，A と B の間の歩行を繰り返すだけのシーケンスが生成される場合がある．そこで，この問題を解決するために，次の歩行先に直前の頂点が選択される確率 p_{back} を与える．

4.3.2 流体シミュレーション結果画像の CNN による推定

画像ベース手法では，CFD シミュレーション結果の画像をもとに CNN を用いたポイ捨て分布を推定する．CNN はニューラルネットワーク手法の一種であり，画像認識や音声認識などにおいて高精度を誇る．特に，画像認識の国際コンペティション ImageNet Large Scale Visual Recognition Competition (ILSVRC) において 2012 年に CNN を用いた AlexNet [35] が 2 位以下に大差をつけて優勝したことを機に，それ以降画像認識においては CNN が広く用いられている．この背景を受けて，本研究でも画像データと CNN を用いてポイ捨てゴミの回帰推定を行った．

まず，CFD シミュレーションの結果を GIS 上でヒートマップとして表示する．図 4.13 は両国駅を中心に風向を 15 度として，風速 5m/s の風を発生させた時のシミュレーション結

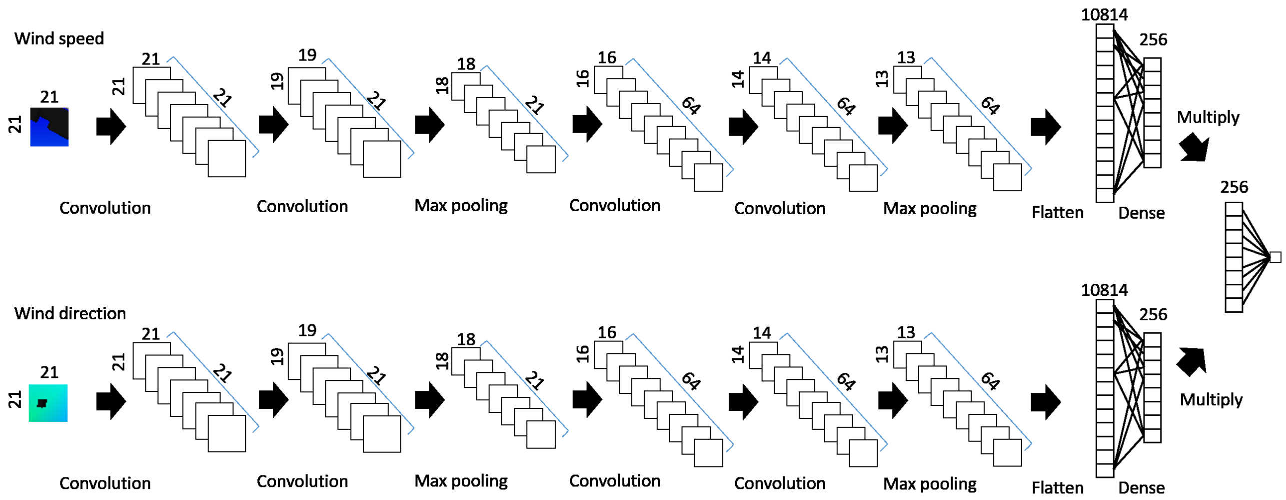

図 4.14: ゴミの個数の予測に使用する CNN のモデル

果から，平均風速をヒートマップ表示させたものである．青が平均風速が遅く，赤に近づくほど速い．黒が建物を，灰色が建物や大型施設の敷地を表している．ライトグレーは道路縁のポリラインをポリゴン化したものである．ポイ捨てゴミはこれらの領域内には存在しない．また，建物に POI 情報がある場合は建物の重心点に POI の種類ごとにマークを表示している．同様に平均風向についてもヒートマップ画像を生成する．この画像を Graph のときと同条件で格子分割し，各画像にゴミの個数の正解ラベルを付与して CNN の学習を行う．図 4.14 に CNN モデルを示す．平均風速の画像と平均風向の画像を異なるユニットで学習し，最終的に乗算する．

4.3.3 CNN と Graph Embedding を組み合わせた推定

近年，Knowledge Graph Completion において，ナレッジグラフのトリプルだけでなく，ナレッジグラフ内に含まれている画像やテキスト情報も利用することで，高い精度を記録することが報告されている．Xie ら [68] はナレッジグラフのトリプルだけでなく，ナレッジグラフ内のエンティティに関する説明文も利用して学習する Description-Embodied Knowledge Representation Learning (DKRL) を提案している．また，Xie ら [69] は，ナレッジグラフのトリプルだけでなく，トリプルに関する画像も使用して学習する Image-embodied Knowledge

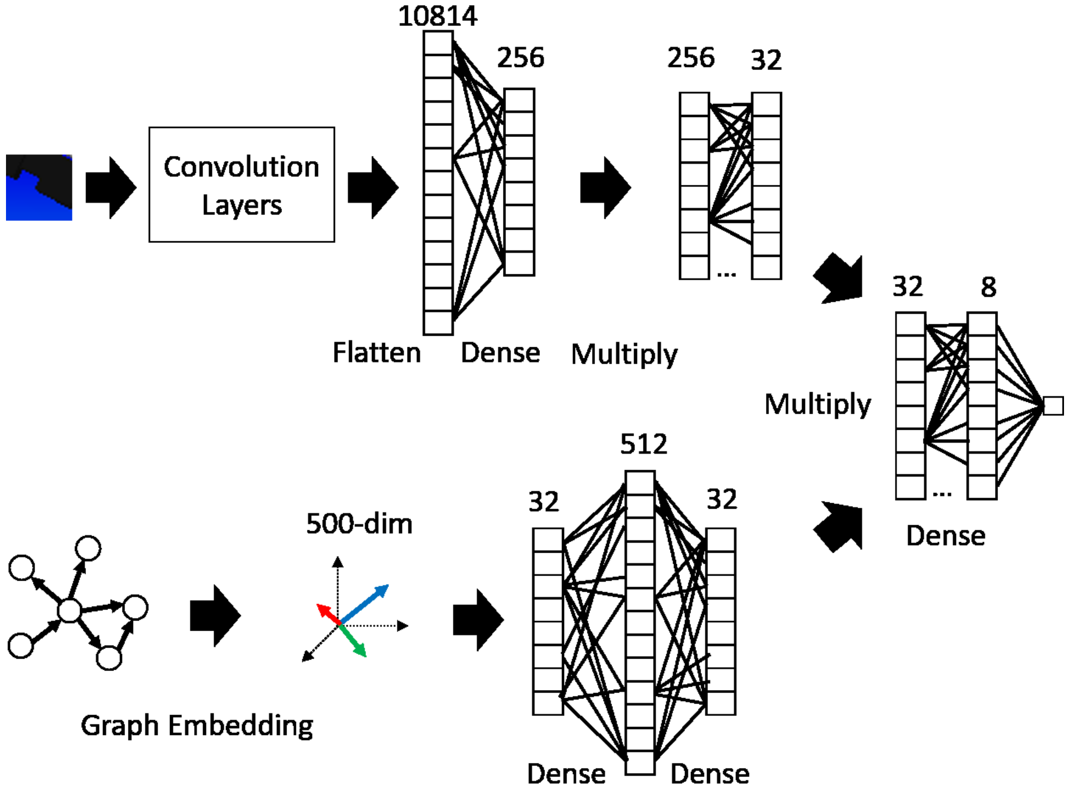

図 4.15: ゴミの個数の予測に使用する画像-グラフ結合手法のモデル

Representation Learning (IKRL) を提案している．DKRL と IKRL は大規模ナレッジグラフを対象とした Translation-based モデルであり，TransE，TransR などと同じ系統であることから本研究への直接採用は適していない．しかし，これらの研究の背景を受けて，4.3.1 節のグラフベース手法と 4.3.2 節の画像ベース手法を組み合わせることにより，本研究においてもさらなる精度向上が見込めると考えた．そこで，ニューラルネットワークモデル上で，グラフベース手法と画像ベース手法を結合した，図 4.15 のモデルを用いてポイ捨てゴミの推定実験を行った．

グラフベース手法の部分では Graph Embedding に RDF2vec を使用し，生成した 500 次元のベクトルデータを入力として 32→512→32 ユニットからなる 3 層の全結合ニューラルネットワークを使用して学習する．図 4.12 のように三層目から出力層へは直接繋がず，画像ベース手法の Convolution 層の後の全結合ニューラルネットワーク部分と結合する．

表 4.6: ポイ捨てゴミの個数の推定手法の精度比較

<table>
<tr><th colspan="3">手法</th><th>RMSE</th></tr>
<tr><td rowspan="5">グラフベース手法</td><td rowspan="2">Translation-based model</td><td>TransE</td><td>3.22</td></tr>
<tr><td>TransR</td><td>4.52</td></tr>
<tr><td rowspan="3">Graph walk-based model</td><td>node2vec</td><td>3.01</td></tr>
<tr><td>RDF2vec</td><td>1.62</td></tr>
<tr><td>ExtendedRDF2vec</td><td>1.50</td></tr>
<tr><td>画像ベース手法</td><td colspan="2">CNN</td><td>1.47</td></tr>
<tr><td>画像-グラフ結合手法</td><td colspan="2">CNN+ExtendedRDF2vec</td><td>0.24</td></tr>
</table>

画像ベース手法の部分では図 4.14 のように 256 ユニットの層から直接出力層へは繋がず，32 ユニットに圧縮してグラフベース手法とノード数を揃えている．32 ユニットの状態で 2 つのモデルを結合し，最後に 8 ユニットの層挟んで出力をする．いずれも活性化関数は Rectified Linear Unit (ReLU) を使用し，最適化手法には RMSprop [63]，Adaptive Gradient Algorithm (AdaGrad) [18]，AdaDelta [70]，Adaptive moment estimation (Adam) [34] を試し，最も精度の高くなった Adam を使用している．

4.3.4 評価

本節ではグラフベース手法と画像ベース手法および両手法の結合手法よるゴミのポイ捨て推定実験を行い，手法を評価する．

グラフベース手法の評価

本実験では両国駅，有楽町駅，新横浜駅の 3 地域分のデータを使用し，構築したナレッジグラフのトリプル数は 169,614 であった．このナレッジグラフから各 Cell リソースが持つベクトル表現を取得するため，本研究では TransE，TransR，node2vec，RDF2vec，改良した RDF2vec（ExtendedRDF2vec）で実験を行った．これらの手法で Graph Embedding を行い，図 4.12 のモデルのニューラルネットワークを用いて，10 分割交差検証による推定実験を行った結果を表 4.6 に示す．RDF2vec において，深さ 3, WLRDF の反復回数は 7 回,

図 4.16: グラフベース手法の学習過程

各インスタンスにおけるグラフウォークの試行回数は 500 まで，Skip-Gram アルゴリズムの window 幅は 5, 出力ベクトルの次元数は 500 としている．また，ExtendedRDF2vec では p_{back} の値を 0 から 1 の間で変更して実験し，最も精度の高くなった $p_{back} = 0.8$ としている．すべての手法でエポック数は 100 としている．精度の評価指標には次式の Root Mean Square Error (RMSE) を使用する．

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(x_i - x_i')} \tag{4.9}$$

N は予測データ数, x_i は予測値, $x_{i'}$ は正解値である．

結果として，Graph Embedding を用いる手法としては我々の提案手法である ExtendedRDF2vec が最も精度が高くなった．

図 4.16 に，グラフベース手法において ExtendedRDF2vec を使用したときの 100 エポックの学習過程を示す．描画結果を見やすくするため精度指標には RMSE を二乗した Mean Square Error (MSE) を使用しており，青色の線が訓練データにおける MSE，橙色の線がテ

図 4.17: 画像ベース手法の学習過程

ストデータにおける MSE である．この結果から訓練データの MSE がテストデータの MSE に比べて極端に良くなっていることが分かり，グラフベース手法は過学習を起こしている可能性が考えられる．過学習の回避のために L1L2 正則化や Dropout 層の追加を行って実験をしたが，ほぼ効果は見られなかった．しかし，表 4.6 は 10 分割交差検証の結果であり，テストデータに関する RMSE が 1.50 と高精度であることが撤回されるわけではない．したがって，今後データ数を増加させることで過学習が抑制され，グラフベース手法のさらなる精度向上が期待できると考えられる．

画像ベース手法の評価

グラフベース手法と同じく両国駅，有楽町駅，新横浜駅の範囲で，平均風速と平均風向のヒートマップ画像を使用して，図 4.14 の CNN モデルで 10 分割交差検証による推定実験を行った．エポック数はグラフベース手法と同じく 100 とした．結果として表 4.6 の通り

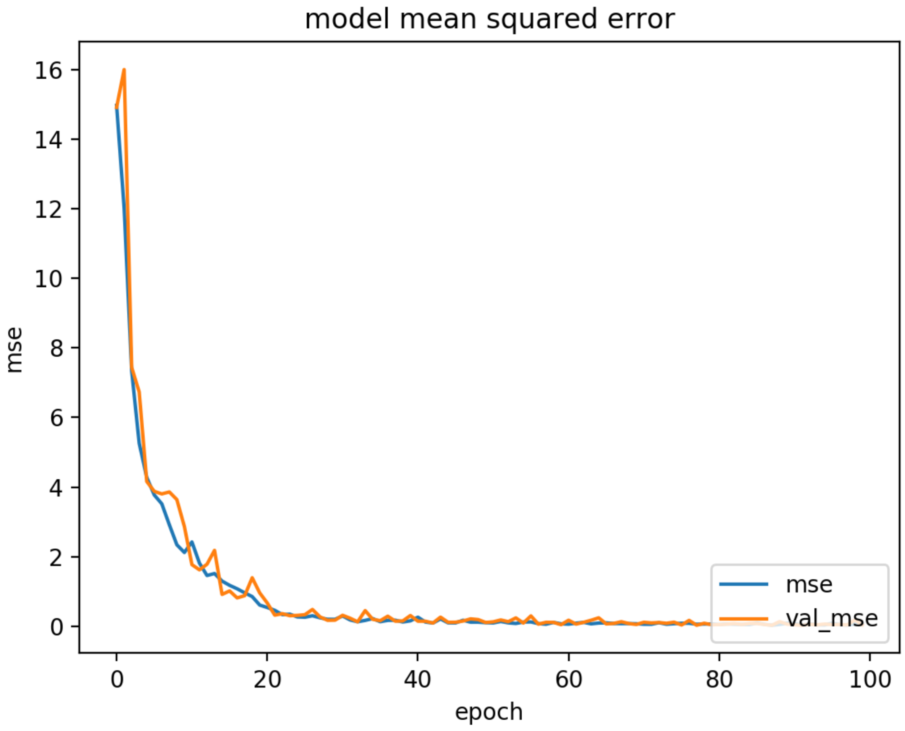

図 4.18: 画像-グラフ結合手法の学習過程

RMSE は 1.47 となり，グラフベース手法より若干精度が良くなったがほぼ同程度であった．

図 4.17 に画像ベース手法における 100 エポックの学習過程を示す．この結果から，画像ベース手法では訓練データの MSE とテストデータの MSE のずれが少なく，過学習が起きていないことが確認できる．また，図 4.16 と比較すると，ニューラルネットワークモデルにおいて，グラフベース手法は画像ベース手法より学習の収束速度が早いことがわかる．

グラフベースと画像ベースの結合手法の評価

図 4.15 のモデルにおいて RDF2vec の 10 分割交差検証での実験を行った結果，表 4.6 の通り RMSE は 0.24 となり本研究において最高精度を記録した．この時，CNN に関しては図 4.14 と同じであり，Graph Embedding に関しては ExtendedRDF2vec を使用した．ExtendedRDF2vec のパラメータは，$p_{back} = 0.6$，歩行回数最大 500，深さ 3，WLRDF の反復回数 6 回，window 幅 5，ベクトル次元数 500 である．図 4.18 に 100 エポックの学習過

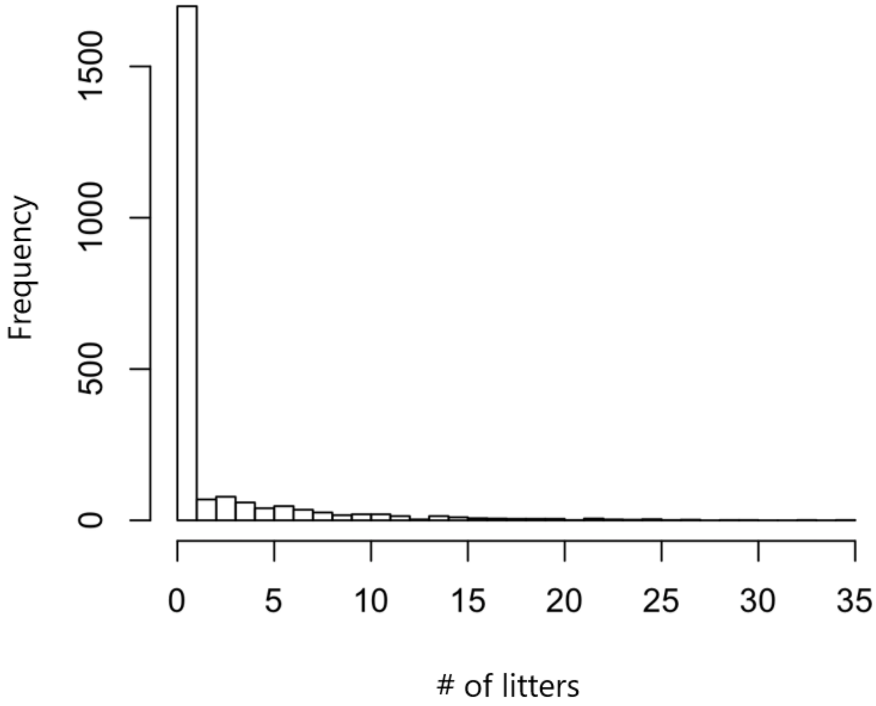

図 4.19: セル内のゴミの個数の分布

程を示す．図 4.18 から 20 エポックを過ぎた辺りという早い段階で学習できていることや，訓練データとテストデータの MSE に大きな差がないことから過学習を起こしていないことが確認できる．図 4.16，図 4.17 と合わせて比較してみると，グラフベース手法の収束速度が速いという性質が生きたまま，データ数の増加により過学習が抑制されたと見ることができる．

データの性質を考慮した評価

本研究で使用したデータはポイ捨て調査を行った道路上のゴミのデータであり，調査を行っていない道路に関してはゴミのデータが無い．グリッド分割の性質上，駅を中心とした矩形選択となるため，グリッド分割した各セルのうち調査対象の道路を含んでいないセルが存在し，このセル内のゴミの個数は 0 となっている．また，グリッド分割したセルの中には建物と被っているセルも多く存在し，これらのセルに関してもゴミの個数は 0 である．図 4.19 に両国駅，有楽町駅，新横浜駅における 1 セル内のゴミの個数の分布を示す．デー

表 4.7: データの偏りを考慮した精度比較

データ	推定手法	RMSE
全データ	CNN+ExtendedRDF2vec	0.24
	すべてを最頻値 (0) と推定	4.52
値 0 のデータを除外	CNN+ExtendedRDF2vec	0.34
	すべてを最頻値 (3) と推定	6.92

タ数は 2202，平均値は 1.73，中央値は 0，最大値は 35，最頻値は 0 である．この図からわかる通りゴミの個数が 0 であるセルが極端に多く，その殆どは調査を行っていない道路に関するセルと建物と被っているセルである．したがって，本研究で使用したデータセットは偏りの激しいデータであり，このようなデータセットを使用した場合は最も頻度の多い観測値に推定することで高精度を記録してしまうという問題がある．例えば，本研究ではすべてのセルにおいてゴミの個数を 0 として推定すれば一定の精度を達成することが予想できる．そこで，データの偏りを考慮してた精度比較の結果を表 4.7 に示す．全テータを使用した場合，セル内のゴミの個数をすべて 0 として予測した場合の RMSE を算出したところ $RMSE = 4.52$ となった．さらに，すべて 0 と推定した場合の精度と本研究の画像-グラフ結合手法の精度 ($RMSE = 0.24$) に有意な差があるかウィルコクソンの符号順位検定を行ったところ，サンプルの予測データ数 220，p 値 <2.2e-16 となり有意水準 0.01 を下回り有意に差があることが確認できた．したがって，データに偏りがあることを考えても本研究の提案手法は有意に高精度で空間的な欠損値を推定できたと言える．

また，実際のユースケースにおいては建物と被っているセルや，調査範囲外の道路に関するセルを予測する必要は無いと考える．そこで，ゴミの個数が 0 であるセルをすべて除いたデータを使用して推定実験を行った．このときのデータの分布を図 4.20 に示す．ゴミの個数が 0 のセルを除いたデータ数は 571，平均値は 6.67，中央値は 5，最頻値は 3 である．このデータを使用して画像-グラフ結合手法の 10 分割交差検証での推定実験を行った結果，表 4.7 に示す通り，$RMSE = 0.34$ となりデータの偏りが大きい場合とほぼ変わらずに高精度を記録した．また，すべて最頻値の 3 と推定した場合は $RMSE = 6.92$ となり，画像-グラフ結合手法と差があるかウィルコクソンの符号順位検定を行ったところ，サンプルの

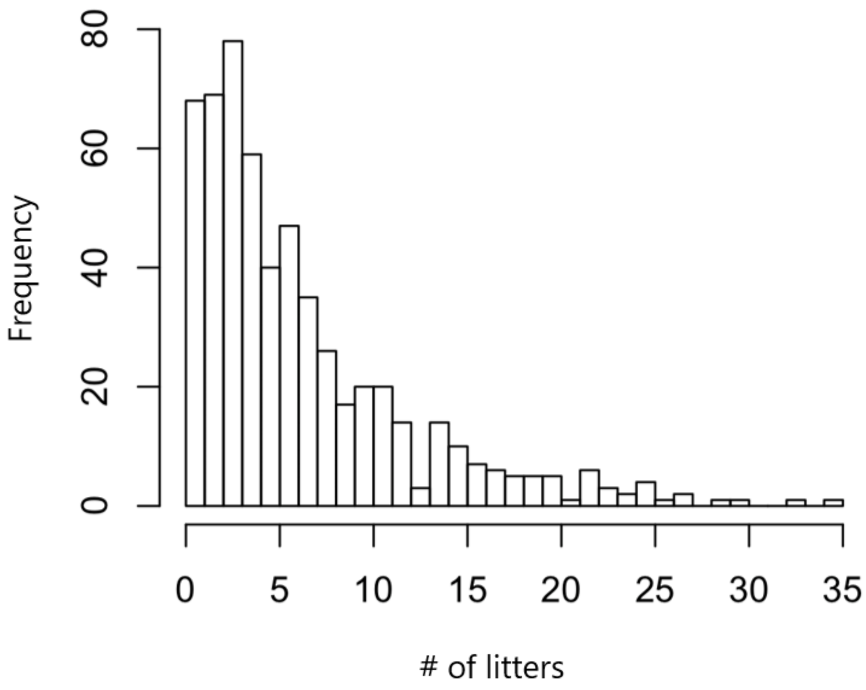

図 4.20: セル内のゴミの個数の分布（0 を抜いた場合）

予測データ数 57，p 値 <5.134e-11 となり有意水準 0.01 を下回り有意に差があることが確認できた．したがって，実用シーンを想定してデータの極端な偏りを除いた場合でも，本研究の提案手法は有意に高精度で空間的な欠損値を推定できたと言える．

4.3.5 考察

グラフベース手法の考察

node2vec は DeepWalk と LINE の改良版として高い精度を誇り，Graph Embedding においては強力な手法の一つであるが，表 4.6 に示したとおり本研究においては RDF2vec が上回った．これは，RDF2vec がエッジの向きを考慮してウォークしている点や，ウォークパスにエッジラベルを含んでいる点が影響していると考えられる．一方で Translation-based モデルの TransE, TransR の精度が悪い理由は，埋め込み表現の学習手法が RDF2vec や node2vec などの Graph Walk-based モデルと大きく異なる点にあると考えられる．本研究

ではポイ捨てゴミの予測をセル単位で行うため，ナレッジグラフ上で似ているセル同士はベクトル空間上で近くに位置するように学習できることが好ましい．TransE はトリプル $(head, relation, tail)$ における $head$ のベクトルと $relation$ のベクトルの和が $tail$ とベクトルに近づくように学習する．そのため，Link prediction 問題によく使用され，DBpedia や Freebase などのノードとエッジの種類が多い大規模なナレッジグラフの分類問題では良い結果を得られる．しかしながら，本研究のように地理空間かつグリッドベースであり，エッジとリソースの種類が少ないナレッジグラフにおいては適していない．したがって，ノードの近傍情報をうまくベクトル空間で表現できる RDF2vec や node2vec のほうが精度が高くなったと考えられる．

本研究で使用した気流シミュレーションは，人の流れと都市空間の構造的特徴を反映し，都市問題が発生しやすい地点を特徴づけることができる．実際に，気流シミュレーションを用いずに地理空間情報のみを使用して推定を行った場合，RMSE は 1.71 となり精度は下がる結果となった．従って，気流シミュレーションは本手法では重要な役割を果たしている．

本手法における推定精度に関わる要因としてはナレッジグラフの構造（スキーマ），ベクトル空間への埋め込み手法，回帰予測手法の大きく分けて 3 つの要素があり，推定精度の良し悪しは必ずしもナレッジグラフのスキーマの良し悪しや枠組み全体の評価には繋がらない．我々の知る限り本手法のように地理空間情報をグリッド分割した上でナレッジグラフ化する既存手法は存在しておらず，ベースとなる既存オントロジーも存在していない．したがって，既存オントロジーに基づいたスキーマの設計ではなく，本研究で新たに提案したスキーマ設計となっている．このスキーマ内ではクラス制約やプロパティの domain と range の制約に違反しておらず RDF としての適切性は充分である言える．実験対象とする都市によって都市空間構造は大きく異なるため，その都度，どの粒度まで知識を記述する必要があるかということについては検討の余地がある．しかし，セル単位で都市問題を予測する場合においてはセル単位でベクトル表現を得る必要があり，建物，道路，POI，気流シミュレーション結果（人の流れ）という大きく 4 つの種類の知識を記述し，これらの知識をセルのインスタンスから紐付けるという構造は変わらないと予想でき，スキーマの基本構造は妥当であると考える．

画像ベース手法の考察

グラフベース手法より若干精度が向上した要因として，CFD シミュレーション結果のヒートマップ表示機能が考えられる．CFD シミュレーションの観測値は格子点に割り当てられているため，一つのセルが持つ観測値は 4 つである．しかし，Airflow Analyst による CFD シミュレーション結果のヒートマップ表示機能を使用すると，4 つの観測値をもとにグラデーション表示が可能である．これにより，格子点単位ではなく 1px あたりに観測値の近似値を保持することができる．本研究では 1 セルあたり 21×21px の画像であり，この画像を 3×3 のフィルタで畳み込んでいる．したがって，グラフベース手法より CFD シミュレーション結果の特徴を詳細に学習できたのではないかと考えられる．

また，画像の場合はピクセル単位で建物の形状や道幅を捉える事ができるため，より地理空間的な特徴を学習できた可能性がある．画像ベース手法では各セルごとに 21×21 ピクセルの画像に，平均風速（または平均風向），建物，建物の POI，道路縁，道路などのすべての情報が含まれている．さらに，それぞれのオブジェクトの違いを明確にするために色を分けている．特に，平均風速については赤，緑，青の 3 色のグラデーションとして表現し，平均風向については角度であることから 0 度と 359 度に差が出ないように，虹の七色でグラデーション表示している．建物については，平均風速，平均風向と色が被ることの無いように，黒や灰色などグレースケールを用いている．これにより，グラフベース手法に比べて数値情報に関する表現の幅は広く，より数値的な特徴を学習できたと考える．また，セル内における建物の形状や位置，道路形状，それらから分かる交差点のパターンなどについても学習できていたのではないかと考えられる．本研究では道路幅のデータを取得することができず，グラフベース手法では建物形状や道路幅を学習させていないため，これが精度差として現れたと考えられる．一方で，グラフベース手法ではこれらのデータを学習していないにもかかわらず，画像ベース手法と大きな差がつかなかったことは新たな発見であった．グラフベース手法では道路幅の代替情報として，ナレッジグラフに道路の種類が含まれている．また，建物と道路との隣接情報を記述している．これが，グラフベース手法において道路に関するセマンティクスを学習し，画像ベース手法における道路の学習のような働きを見せたのではないかと考える．今後グラフベース手法では，建物や道路のポリゴン情報をもとにグラフ化することで，さらなる精度の向上が期待できる．

また，画像ベース手法では平均風速と平均風向の画像を別々に学習させ，最終的に乗算するというモデル構成になっている．一枚の画像に複数種類の観測値を含めることができないため，このようなアプローチとした．一方，グラフベース手法ではナレッジグラフに複数種類の観測値を含める事ができるため，学習を分けること無く単一入力からなるニューラルネットワークモデルとなった．そのため，学習に用いるデータ量としては画像ベース手法が多くなり，精度向上につながったのではないかと考えられる．

グラフベースと画像ベースの結合手法の考察

グラフベースと画像ベースの結合手法が大幅に精度を伸ばした要因は，グラフベースと画像ベースのそれぞれがCFD結果および地理空間情報に関して異なる特徴を学習し，情報量が増えたためであると考える．4.3.4節で述べたように，画像ベース手法では数値情報に関する表現の幅が広く，より数値的な特徴を学習できたと考える．一方，グラフベース手法に関しては，数値情報に関しては等間隔分類しているため画像に比べて表現の幅が狭くなり，微小な特徴を学習できていない可能性が考えられる．また，元リテラル値のリソースはプロパティを持たないため，グラフウォーク時に必ず終着点となりシーケンス集合内で出現頻度は少なくなる．そこで，元リテラル値を含むウォークパスを複製し，数値情報に関する特徴をより多く学習させる実験を行ったが，精度の向上は見られなかった．このことからも，数値情報に関してはグラフベース手法よりも画像ベース手法が特徴を学習し，結合手法においての精度向上に貢献していると考える．しかし，次元削減のようにある程度情報量を落とすことが精度向上につながる可能性もあるため，グラフベース手法における数値情報の分類に関してはさらなる検証が必要である．

グラフベース手法では画像ベース手法には含まれていない，隣接セルの関係性やオブジェクト間の関係性などのセマンティクスを学習している．特に，飲食店の集中するエリアでは，複数のセルにまたがって多くのゴミがポイ捨てされていることがある．このようなポイ捨ての現状を考えると，各セルにおける特徴だけでなく，隣接セルも考慮した一定の範囲の特徴を学習したことがグラフベース手法の精度向上に貢献し，結合手法においての精度向上にも貢献したと考えられる．同様に，グラフでは建物や道路のpartOf関係やPOIとの関係，建物と道路の隣接関係などを明示的に記しており，画像よりもこれらの意味的な

特徴を学習できたと考えられる．

このように，画像ベース手法とグラフベース手法では，それぞれで正確に学習できた特徴が異なっているため，結合することでより多くの特徴を正確に学習でき最高精度を記録できたと考える．今後，画像ベース手法とグラフベース手法のそれぞれで，どの要素を学習させることが精度向上につながるか調査することが過大である．これにより，都市の地理空間的な特徴と人の行動に関係の深い都市問題に関して，空間的欠損値推定手法の最適なモデルの設計が可能になる．

本手法の適用対象としてゴミのポイ捨て問題を例とした．より一般化すると，都市の地理空間的な特徴と人の行動に関連の深い都市問題のうち，特に発生地点が一定の範囲において広く密に分布している問題において適用が可能である．同様の問題として，例えば渋滞のような問題に本手法が適用できると考える．ただし，気流シミュレーションにおけるパラメータやベクトル化手法およびニューラルネットワークモデルなどは，個々の問題によって調整が必要となる．

4.4 都市問題に関するナレッジグラフの活用例の検討

本章では，本研究で構築した時空間的に詳細な都市問題ごとのナレッジグラフの活用例について説明する．

4.4.1 SPARQLを用いたナレッジグラフの活用例

SPARQLを用いたナレッジグラフの活用例として，次のSPARQLクエリを用いてSPARQLエンドポイントから時間帯別の放置自転車台数の情報を取得して可視化したところ，図4.21の結果がえられた．

```
@prefix event: <http://purl.org/NET/c4dm/event.owl#>
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?hour (AVG(?value) AS ?average)
WHERE {
        ?s rdf:value ?value .
        ?s event:time/time:hour ?hour .
} GROUP BY ?hour ORDER BY ASC(?hour)
```

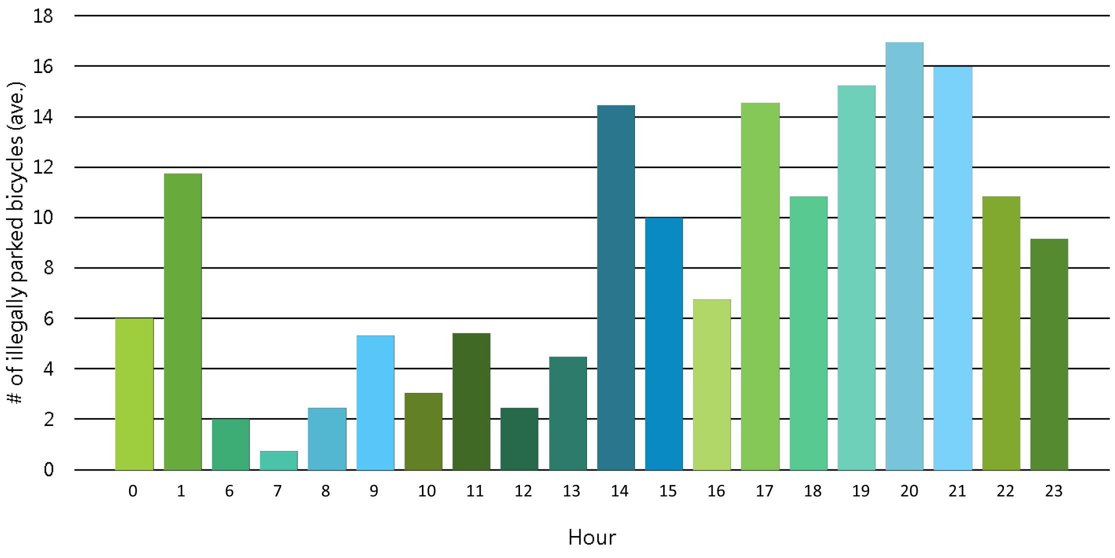

図 4.21: 全観測地点における一時間ごとの平均放置自転車台数

この結果から，朝から昼の時間帯より夕方から深夜の時間帯の方が放置自転車が多いことがわかった．一般的に通勤・通学の時間帯に放置自転車が多いと認知されているが，本研究の実験地域においては逆の結果が示された．このような結果は，本研究のように時系列的に詳細度の高いスキーマ設計にすることで，SPARQL により容易に取得することができる．次に，以下の SPARQL により得られた，近くの POI の統計情報を表 4.8 に示す．

```
SELECT DISTINCT ?ntype count(?ntype) WHERE {
  ?place a geo:SpatialThing .
  ?place gn:nearby ?near .
  ?near a ?ntype ;
} GROUP BY ?ntype
ORDER BY DESC(COUNT(?ntype))
```

この結果，本研究で得られた放置自転車の観測地点の近くには飲食店，小売店，食料品店等が多いことが分かった．本研究により，このような課題解決に役立つ可能性のある検索を SPARQL を用いて簡単に行うことが可能になった．さらに，放置自転車 LOD のように地理情報を詳細に含めることで様々な LOD とつながる可能性があり，複雑な SPARQL クエリによる問い合わせやオントロジー推論により，従来では考えられなかった問題の原

表 4.8: 放置自転車観測地点の近くの POI トップ 10

POI types	Amount of POI types
Restaurant	473
Retail	380
Food	366
Shop	365
Building	310
Health	116
Hairdresser	80
Hospital	63
Financial	59
Bar	58

因特定に役立つ．例えば，LinkedGeoData の is-a 関係に基づく推論により POI の上位概念の持つプロパティを使用した検索や，`owl:sameAs` や `owl:equivalentClass` 関係に基づく推論により，同等クラスとして付与されるプロパティを使用した検索などが可能である．また，表 4.1 の統計情報も，放置自転車 LOD が地理的に詳細度の高いスキーマ設計になっているため，同様に SPARQL により容易に取得することが可能であり，放置自転車 LOD のスキーマ設計は実用面において一定の有用性があると考えられる．しかし，本研究で提案したスキーマでは既存オントロジーの一部のクラスの制約を順守できていない部分も存在する．また，`event:factor` プロパティの値の粒度が異なるという問題も生じる。オントロジーの意味的整合性を順守しすぎるあまりに構造が複雑化してしまうことは，ナレッジグラフの利便性を低下させるおそれがある．今後はオントロジーの制約を順守しつつ実用性の高いスキーマを設計するため，新たなオントロジーの設計が必要になると考える．

4.4.2　ナレッジグラフを活用した可視化アプリケーション

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なる．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

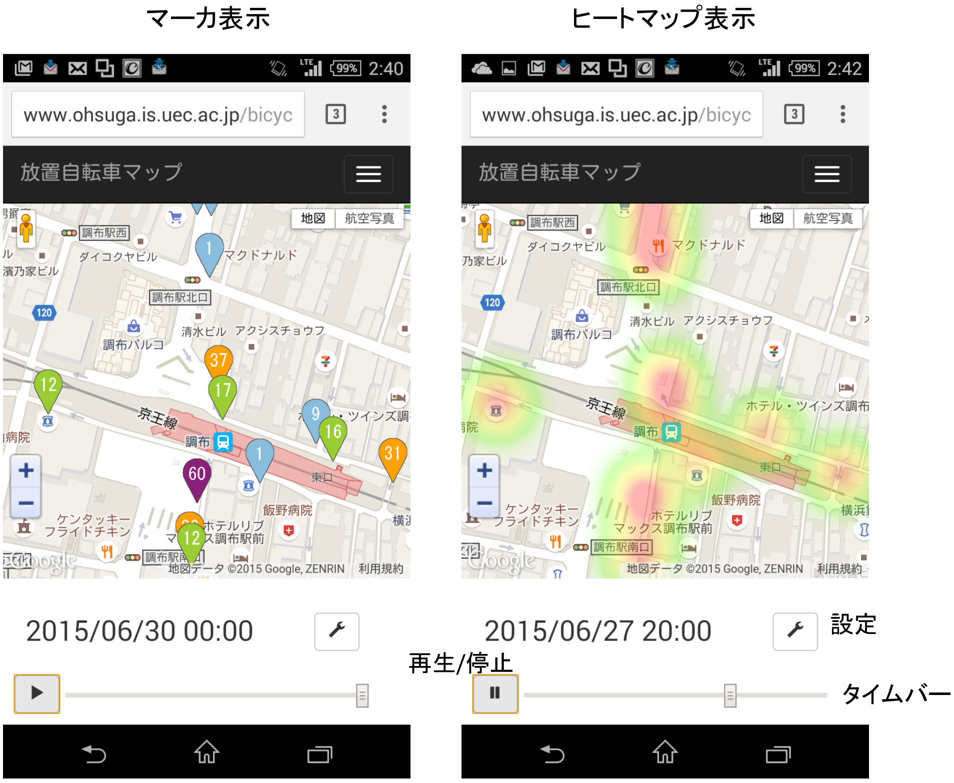

図 4.22: 可視化したアプリケーションの画面例

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 の活用の一例として，放置自転車に時系列変化を可視化する Web アプリケーション[7]を公開した．アプリケーションの動作例を図 4.22 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カを選択し，再生ボタンを押すと自転車の放置状況がアニメーション表示される．図 4.22 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カでアニメーション表示させた画面である．アプリケーションのバックグラウンドでは下記の SPARQL が SPARQL エンドポイントに送信されており，可視化に必要なデータを放置自転車 LOD から取得している．

```
@prefix ipb: <http://www.ohsuga.is.uec.ac.jp/bicycle/resource/>
```

[7] http://www.ohsuga.is.uec.ac.jp/bicycle/

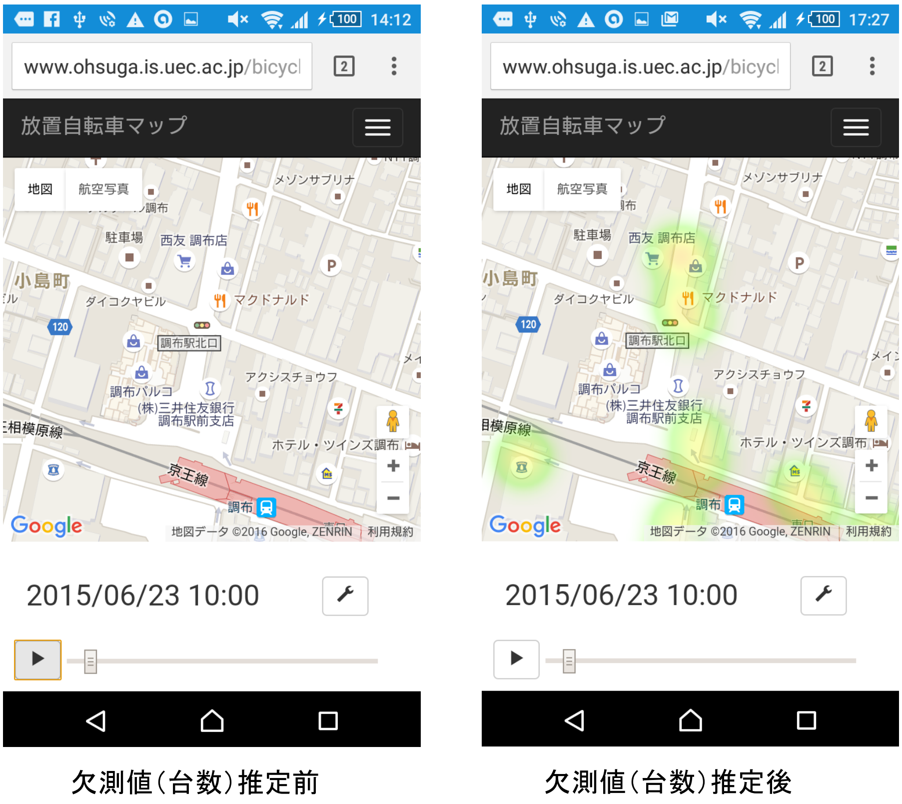

図 4.23: 可視化したアプリケーションの画面例

```
@prefix event: <http://purl.org/NET/c4dm/event.owl#>
@prefix dcterms: <http://purl.org/dc/terms/>
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>
select ?s ?lat ?lon ?num ?date where {
?s a ipb:IllegallyParkedBicycles;
event:place ?place;
event:time [dcterms:created ?date] .
?place geo:lat ?lat;
geo:long ?lon .
optional {?s rdf:value ?num}
optional {?s <http://www.ohsuga.is.uec.ac.jp/ipblod/vocabulary#
    estimatedValue> [rdf:value ?num]}
filter(?date >= "2015-06-23"^^xsd:dateTime && ?date <= "2015-06-30"^^
    xsd:dateTime)} order by ASC(?date)
```

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

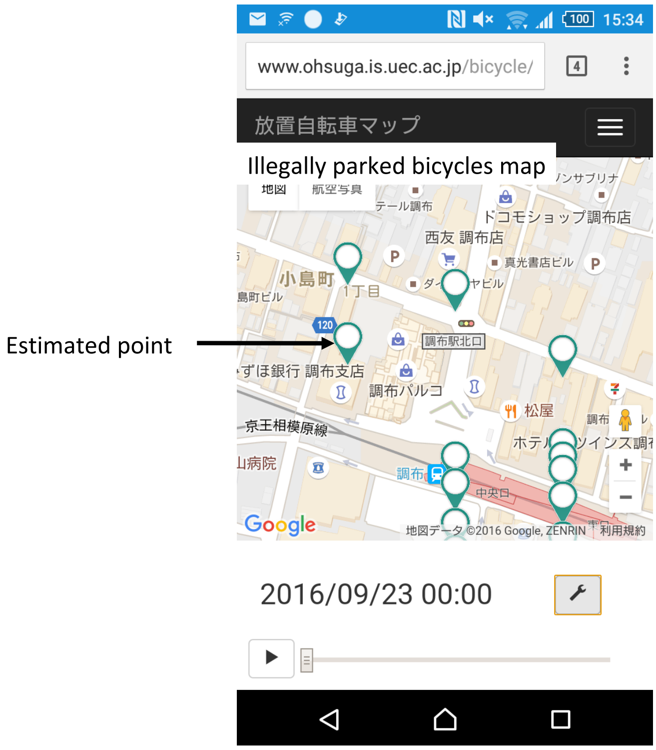

図 4.24: 空間的欠損値の可視化

ネットワークによる欠損推定値も確率とともに放置自転車 LOD に含めており，時系列的に密なナレッジグラフとなっている．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 4.23 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 4.23 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であるのに対し，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．

さらに，空間的欠損値の推定結果を地図にマッピングする機能をアプリケーションに追加した．図 4.24 に推定地点を可視化した結果を示す．この可視化結果はユーザのデータ投稿の手助けとなり，推定地点における新たな観測データの収集が期待できる．

この可視化アプリケーションは，図 3.4 のツイート Web アプリケーションと同一サイト上にあり，ツイート後即時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

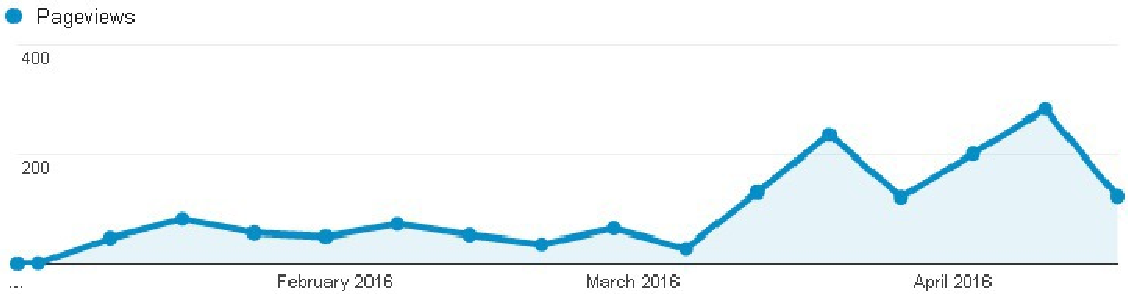

図 4.25: 可視化アプリケーションのページビュー数

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．

可視化活用に関する考察

2016 年 1 月から 2016 年 4 月にかけてアプリケーションのページビュー数を調査したところ，図 4.25 のように増加傾向にあることがわかった．可視化アプリケーションのページビュー数は 2016 年 1 月時点で一ヶ月間で 187 であったが，4 月には一ヶ月間で 705 まで増加していた．また，ユニークユーザ数は 40 人から 118 人に，平均セッション持続時間は 2 分 32 秒であった．セッション持続時間とは，1 ユーザによるウェブサイトの連続的なブラウジング時間である．したがって，可視化アプリケーションの訪問者は公開後から増加し，一度訪れると長く使用していることがわかった．

また，セッション数と位置情報を調査した結果を表 4.9 に示す．集計期間は 2015 年 10 月 1 日から 2016 年 10 月 1 日までの一年間であり，セッション数上位 10 地域を表示している．この結果から，最もデータが多く集まっている東京都調布市より，横浜市，新宿区のセッション数が多いことがわかった．また，調布市，新宿区，横浜市，港区，大阪市，世田谷区ではセッション数が増加していることがわかった．中でも横浜市は約 2 倍，新宿区は約 3 倍にセッション数が増加していることがわかった．さらに 2016 年 4 月までに 68 地域のユーザからのセッションを確認したが，2016 年 10 月までに 118 地域に増加していることがわかった．これらの結果から，放置自転車 LOD および可視化アプリケーションの認知度は徐々に高まっており，本研究を通して放置自転車の問題意識の醸成に貢献できたと考

表 4.9: 市区町村ごとのユーザ数トップ 10

Oct. 2015 - Mar. 2016		Apr. 2016 - Sep. 2016	
City	# of users	City	# of users
Chofu	66	Shinjuku	129
Yokohama	61	Yokohama	127
Minato	48	Chofu	69
Shinjuku	39	Minato	64
Kodaira	35	Osaka	24
Osaka	20	New Delhi	21
Nagoya	19	Chuo	20
Shibuya	9	Fucu	14
Setagaya	8	Setagaya	14
Kawasaki	8	Kodaira	12

えられる．

自治体との連携

本研究では，都市問題の一つとして放置自転車問題解決に向けてナレッジグラフの構築と活用を行ったが，実際に放置自転車の削減に繋がったかどうか確認するためには長期的な評価が必要となる．本研究では，東京都青少年・治安対策本部と連携し，放置自転車LODの活用がどのように問題解決つながるか協議を重ねている．東京都青少年・治安対策本部は放置自転車を深刻な問題と捉えており，駅前放置自転車クリーンキャンペーンを実施している．2015年に行われた第32回クリーンキャンペーンでは，東京都によるポスターの作成と配布，各市区町村によるポスターの作成と配布，新聞の発行，人気タレントを起用した広報動画の作成，街頭スクリーンやデジタルサイネージによる放映，PRティッシュの配布，児童によるポスターコンクールの開催など，多数の策を実施した．2016年度では同様の取り組みに加えて，ポスターに美少女キャラクターを起用したPRなどを行い，毎年力を入れている．さらに，2017年度では本研究との連携を行い，本研究は放置自転車LODの可視化を通じてこのキャンペーンに協力し，問題意識の醸成とデータの二次利用による

問題解決促進に取り組んだ．本キャンペーンのホームページ[8]上で，放置自転車マップとして本研究の取り組みが紹介されている．2016 年度に作成されたポスター数は都作成のものが 44,600 枚，区市作成のものが 406 枚，関係団体作成のものが 26,900 枚であり，駅構内，電車・バス内，学校，関係機関・団体内に掲示された．リーフレットは都作成のものが 390,000 枚，区市町村作成のものが 31,599 枚であった．広報誌・機関紙等は 10,785,770 部発行された．広報・撤去活動参加人員は 13,879 人，撤去台数は 10,652 台であった．このような大規模な社会活動に LOD の研究が加担する事例は他に見ない．今後，東京都青少年・治安対策本部と協議を重ね，問題解決に向けた新たな策を講じる予定であり，本研究の事例が他の都市問題解決に向けた一つのテストケースになると考える．

[8]http://www.seisyounen-chian.metro.tokyo.jp/kotsu/kakusyutaisaku/jitensha/houchi/jitensha-campaign01/index.html

第5章　都市問題の課題横断的なナレッジグラフの構築と分析

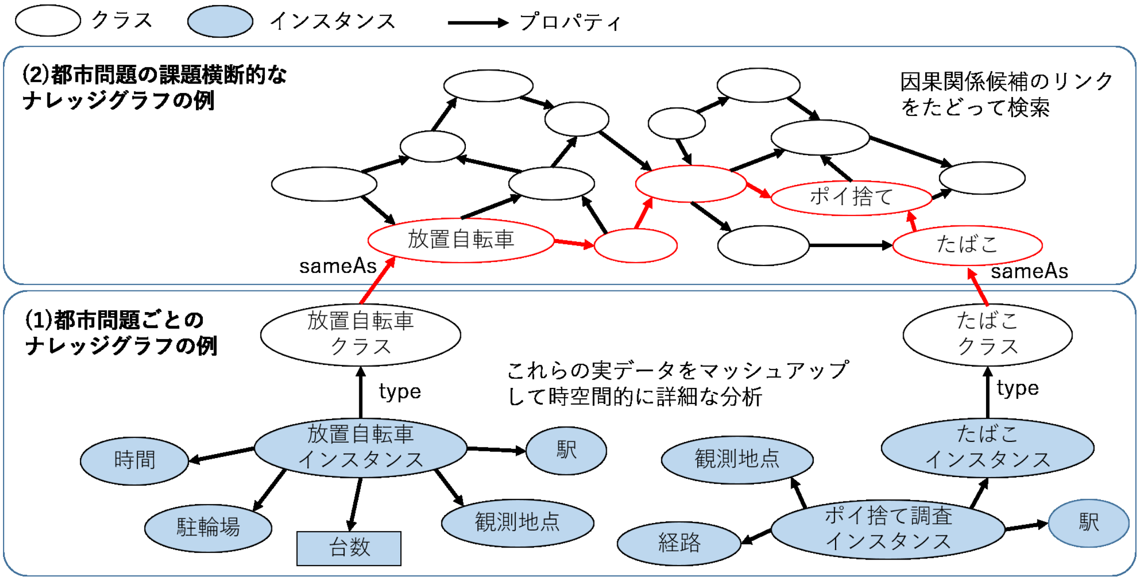

図 5.1: 2 種類のナレッジグラフをまたぐ分析のイメージ図

本章では，複数の都市問題間の因果関係を記述するナレッジグラフの構築と活用について説明する．4 章までで提案した，都市問題ごとに時間的・空間的に詳細なナレッジグラフ (1) の設計，構築，欠損補完および分析活用手法は，一般化の対象範囲として都市の地理空間的な特徴と人の行動に関連の深い都市問題を想定した．一方で，本章で提案する，複数の都市問題間の複雑な関係を記すナレッジグラフ (2) の設計，構築，欠損補完および分析活用手法は，より広義の都市問題全般に適用できる手法である．構築したナレッジグラフの用途としては，(1) は主に個別の都市問題の分析に活用でき，(2) は主に都市問題の因果関係を把握するために活用できる．また，(2) のナレッジグラフは (1) の上位概念的な位置づけであり，(1) のナレッジグラフにおける都市問題名のクラス部分は，(2) のナレッジ

グラフにおける都市問題名のクラス部分と同等である．したがって，図 5.1 に示すように，(1) の都市問題ごとのナレッジグラフから，都市問題名のクラスを辿ることで (2) のナレッジグラフに到達し，(2) のナレッジグラフから都市問題の因果関係リンクを経由して別の都市問題を検索することが可能である．さらに，検索して得られた都市問題について下位概念やインスタンスをたどることで，その都市問題に関する (1) のナレッジグラフに到達し，検索元の都市問題とマッシュアップして時間的・空間的に詳細な分析することが可能である．

5.1 都市問題の因果関係と自治体予算情報を含むナレッジグラフの設計と構築

5.1.1 都市問題の因果関係と自治体予算情報のスキーマ設計

本研究で構築するナレッジグラフの主要な使用用途は，自治体の抱える社会課題（都市問題）の解決施策や予算削減案の導出支援である．具体的には，因果関係を基に都市問題解決策の波及効果や，各都市問題に関する自治体ごとの予算を検索可能にし，都市問題解決施策の作成や予算削減案の作成を支援する．そこで，我々は都市問題の因果関係と自治体の予算情報を記述するスキーマを設計した．図 5.2 に設計したスキーマを示す．図 5.2 の上部は都市問題の因果関係を記述する語彙である．因果関係部分ではリソースは都市問題かそうでないもののどちらかに分類される．因果関係を表すプロパティは大きく分けて `factor` と `affect` があり，推移律を持つ `TransitiveProperty` である．これら二つのプロパティは related プロパティのサブプロパティである．また，クラウドソーシングプロセスにおける一致人数に応じてサブプロパティを定義している．このようにプロパティを定義することで，用途に応じた推論が可能になる．例えば，クラウドソーシングプロセスにおける一致人数の高い因果関係を知りたい場合 `factor_level4` や `affect_level4` プロパティを使用し，一致人数に関係なく因果関係を全て知りたい場合は `factor` と `affect` プロパティを使用する．

図 5.2 の下部は自治体の予算情報を記述する語彙である．自治体の予算情報は Microsoft Excel や Adobe PDF などの表形式データで公開されているため，RDF Data Cube Vocab-

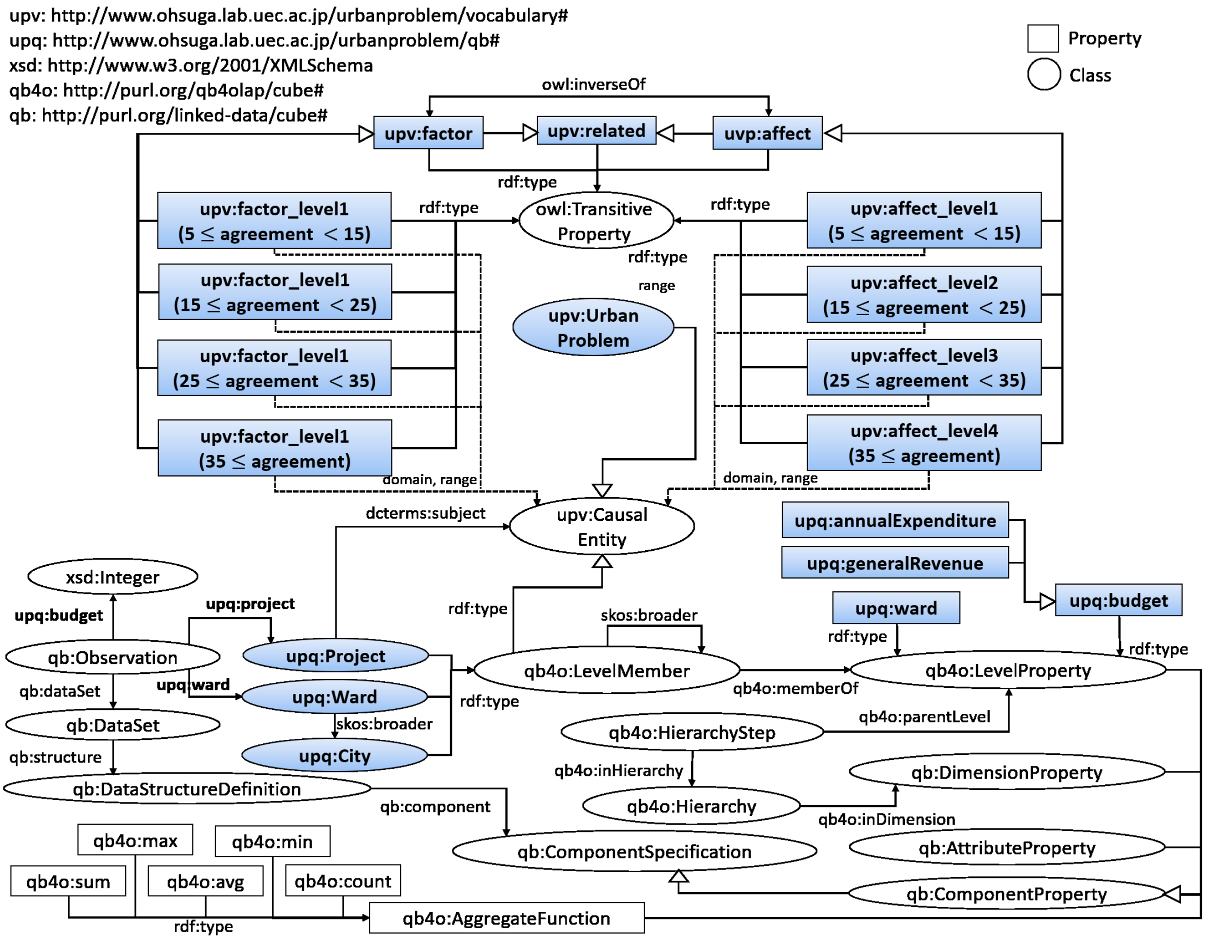

図 5.2: 都市問題の因果関係と予算情報を表す Linked Data スキーマ

ulary を使用して記述する．さらに，本研究では Online analytical processing (OLAP) のデータモデルで使用され，RDF Data Cube Vocabulary の拡張である QB4OLAP[19] に基づいて予算情報を記述する．QB4OLAP により，都市問題対策事業の地域全体の予算合計や部局ごとの予算合計，最大値最小値などを検索することが可能になる．自治体の事業名プロパティの range は事業名クラスであり，事業名クラスは少なくとも一つの dcterm:subject プロパティを持ち，その値は skos:Concept クラスである．このように設計することで，都市問題の波及効果から関連予算まで取得することが可能になる．

5.1.2 データ収集と係り受け解析による因果関係の半自動的な抽出

本研究では，都市問題の因果関係を半自動的に抽出する手法を提案する．提案手法は以下の 4 ステップで構成されている．

(1) 検索エンジンを用いた Web 文書の収集

(2) 係り受け解析による文書からの因果単語候補の抽出

(3) 抽出した因果単語を元にワードクラウドを生成

(4) クラウドソーシングを用いた因果単語候補のフィルタリング

まず，対象とする都市問題名とその同意語を一番目の検索語とし，“要因” の同意語を二番目の検索語として検索エンジンを用いた記事検索を行う．以下，郊外型犯罪の要因を例として説明する．郊外型犯罪の場合，一番目の検索語は “郊外型犯罪” を使用し，二番目の検索語は “要因” とその同意語であり，日本語 WordNet[1]から “要素”，“原因” を取得して使用する．同意語として取得したが都市問題に関する日本語文書において使用頻度の少ないと考えられる，“因子”，“素因”，“導因”，“エレメント”，“ファクター” 等に関しては，これらの検索語がメインとして検索されてしまったため除外した．本研究では検索エンジンとして Google と Bing を使用し，Google Custom Search API[2] と Bing Web Search API[3] に「郊外型犯罪　要因」のようなキーワードパラメータを与えて結果を取得する．取得するファイルタイプとして HTML と PDF を別々に指定し，一番目の検索語と二番目の検索語の全ての組み合わせで記事を 50 件ずつ検索する．HTML と PDF を分けて検索した意図としては，都市問題の因果関係について言及している記事を，行政の調査報告や議事録等と一般人の SNS やブログ等から満遍なく収集するためである．また，記事中における都市問題名の出現回数が低い場合は，この後の要因単語抽出処理でノイズが多く混ざってしまうため，都市問題名の出現回数でフィルタリングを行う．

収集した記事の内，“要因”，“要素”，“原因” を含む文を抽出し，文ごとに形態素解析を行い名詞を抽出する．この後の処理であるクラウドソーシングによる要因単語選択の処理

[1] http://compling.hss.ntu.edu.sg/wnja/
[2] https://developers.google.com/custom-search/?hl=ja
[3] https://azure.microsoft.com/en-us/services/cognitive-services/bing-web-search-api/

において，ある程度意味が通じる名詞となるように，サ変接続名詞の直前・直後に名詞がある場合はそれらを連結して一つの名詞とする．例えば “非行防止” は通常 “非行 (名詞-一般)” と “防止 (名詞-サ変接続)” に分割されてしまうが，本手法ではサ変接続を結合して “非行防止” とする．

次に cabocha [36] を用いた日本語係り受け解析を行い，“要因”，“要素”，“原因” を含む文節と直接的に依存関係のある文節から，前の処理で抽出した名詞句に相当する部分を抽出する．間接的に依存関係のある文節まで取得するとノイズが多くなってしまうため，本研究では直接的に依存関係のある文節のみ使用した．

同様に都市問題の影響に関しての抽出も行う．“影響” の他，日本語 WordNet から “引き起こす”，“惹起”，“もたらす”，“来たす”，“招く”，“誘発”，“生み出す” などの同意語を取得し，これらを二番目の検索語として要因抽出と同様の処理を行う．

5.1.3 クラウドソーシングを用いた因果単語候補のフィルタリング

次に抽出した因果単語を基にワードクラウドを生成し，クラウドソーシングによる因果単語の選択を行う. まず，5.1.2 節で抽出した因果単語の内，類似度の高い単語を文字列の類似度計算結果に基づいて統合した．文字列の類似度計算には Jaro-Winkler 距離 [67] を使用し，類似度が 0.8 以上であった場合に文字列の長い単語に統合した．次に因果単語の出現回数を基にワードクラウドを生成した．ワードクラウドを使用する理由は，ワーカになるべく負担をかけず重要な単語の抽出を容易にするためである．図 5.3 に郊外型犯罪の要因単語のワードクラウドを示す．出現回数が多いほど中心に位置し文字サイズが大きくなる．文字の色はランダムに決定する．このワードクラウドから「郊外型犯罪の原因として考えられるものを 10 個選択する」「郊外型犯罪が引き起こすものや影響範囲として考えられるものを 10 個選択する」というタスクを設定し，クラウドソーシングによる因果単語の選択を行う. 今回は金銭的報酬を与えるマイクロタスク型クラウドソーシングを行った．クラウドソーシングサービスとしてランサーズ[4]を使用し，上記 2 つのタスクの遂行報酬を 50 円として，一つの都市問題につき最大 50 人に作業を依頼した．このうち作業人数の 1 割にあたる 5 人以上に選択された単語を使用する．選択された単語の中に類似の単語がある場

[4]http://www.lancers.jp/

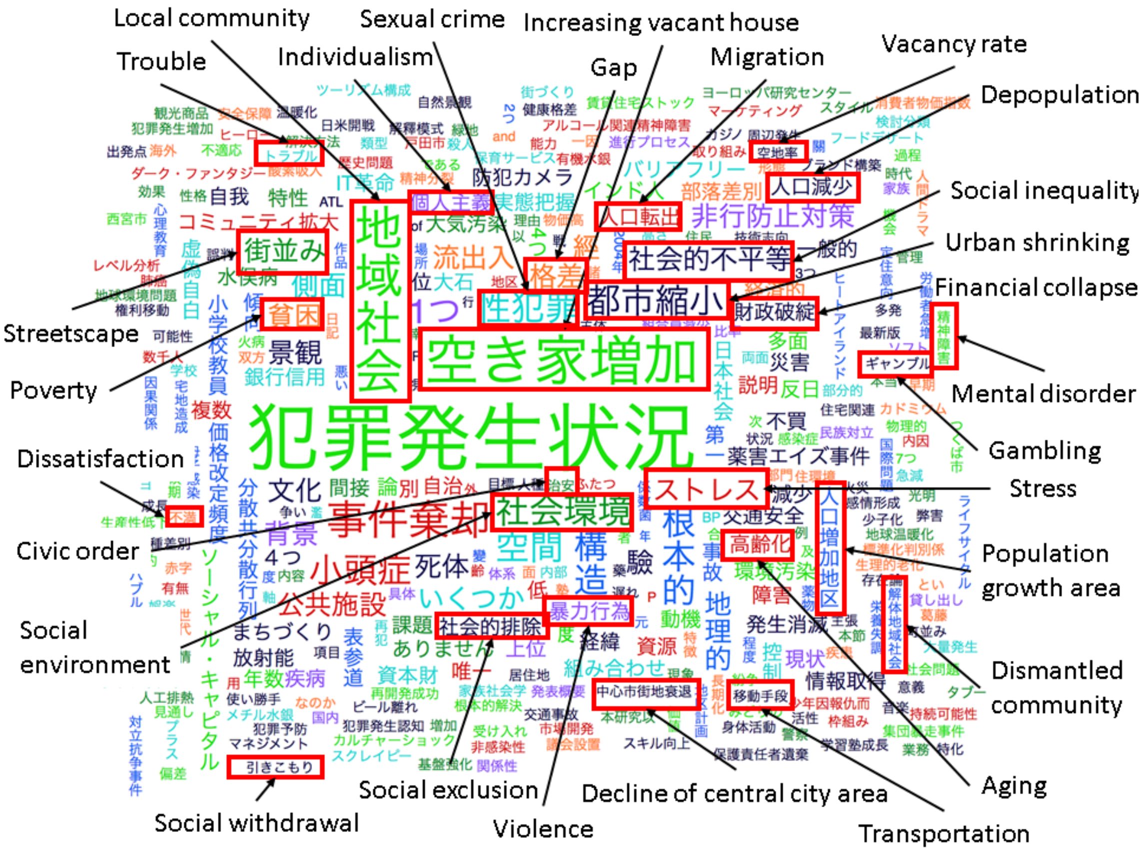

図 5.3: 郊外型犯罪の要因のワードクラウド

合は，選択した人数の多い単語に統一する．図 5.3 内で赤枠で囲まれた単語は実際にクラウドソーシングで 5 人以上に選択された単語である．

さらに，より多くの因果関係を構築するために，クラウドソーシング後に最終的に抽出された因果単語について，提案手法の (1) から (4) のステップを繰り返し実行する．しかしながら，必ずしも全ての単語が因果関係を持つとはかぎらないため，1 回目のループで抽出された因果単語について Google 検索結果の上位 50 件の記事から共起語を調査し，原因，要因，影響，都市などの単語を 336 以上持つ単語 (上位 5%) のみこのステップを実行する．

5.1.4　因果単語抽出結果に基づくナレッジグラフの構築

5.1.1 節で設計したスキーマに基づいて，抽出した因果単語に関するナレッジグラフを Linked Data 形式で構築した．クラウドソーシングの結果はタブ区切りデータで取得でき

るため，このデータを基に Apache Jena[5]を使用して RDF に変換した．具体的には，都市問題は UrbanProblem クラスのサブクラスとして生成され，それ以外は `CausalEntity` のサブクラスとして作成される．各クラス間の因果関係はクラウドソーシングにおける一致人数別に `factor`，`affect` のサブプロパティを用いて生成される．更に，各クラス名を形態素解析して名詞を抽出し，上位クラスとして作成した．また，全てのリソースについて WikiData 内に存在する同名のリソースの `skos:altLabel` の値を取得し，その値を基に代替リソースを生成し，Prov Ontology[6]の `prov:alternateOf` プロパティでリンクする．これにより表記ゆれ問題をある程度解消している．全ての単語は Google Translate API[7]で英語に自動翻訳し，リソースの一意な URI はこの英語を基に生成する．下記に構築される RDF トリプルの例を示す．

```
PREFIX upr: <http://www.ohsuga.lab.uec.ac.jp/urbanproblem/resource/>
PREFIX upv: <http://www.ohsuga.lab.uec.ac.jp/urbanproblem/vocabulary#>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
upr:Temporal_employment a owl:Class ;
rdfs:label ‘‘Temporal employment’’@en , “非正規雇用”@ja ;
rdfs:seeAlso <http://www.wikidata.org/entity/Q667944> ;
rdfs:subClassOf upr:Employment , upv:CausalEntity ;
skos:broader upv:CausalEntity , upr:Employment ;
upv:factor_level1 upr:Homeless ;
prov:alternateOf upr:Part_timer .
upr:Part_timer a owl:Class ;
        rdfs:label ‘‘Part-timer’’@en , “パートタイマー”@ja ;
        rdfs:subClassOf upv:CausalEntity ;
skos:broader upv:CausalEntity ;
        prov:alternateOf upr:Temporal_employment .
upr:Employment a owl:Class ;
        rdfs:label ‘‘Employment’’@en , “雇用”@ja ;
        rdfs:subClassOf upv:CausalEntity ;
        skos:broader upv:CausalEntity ;
        prov:alternateOf upr:Employment_contract .
upr:Employment_contract a owl:Class ;
        rdfs:label ‘‘Employment contract’’@en , “雇用契約”@ja ;
rdfs:subClassOf upv:CausalEntity ;
skos:broader upv:CausalEntity ;
prov:alternateOf upr:Employment .
```

[5]`https://jena.apache.org/`

[6]`https://www.w3.org/TR/prov-o/`

[7]`https://cloud.google.com/translate/docs?hl=ja`

5.1.5 自治体予算情報のナレッジグラフの構築

大阪市は大阪市オープンデータポータルサイト[8]で様々なデータを CC-BY 4.0 ライセンスで公開している．本研究では大阪市をモデルとし，このポータルサイトから予算情報のデータを取得して RDF に変換する．まず，予算情報のデータは表形式の PDF データとして公開されているため，Adobe Acrobat を使用して Excel データ (XLS) に変換する．次に Apache POI[9]を用いてセルから情報を抽出し，Apache Jena を用いて 5.1.1 章で設計したスキーマに基づいて RDF を生成する．

次に，自治体の事業リソースと都市問題および都市問題の因果リソースとのリンク付けを行う．元のデータには事業に関する説明は記述されていないため，事業名称に含まれている単語を基に因果リソースへのリンク付けを行う．しかしながら，「路上喫煙対策事業」と「タバコ」のような関係は既存のシソーラス等を使って取得することはできない．そこで，自治体事業名からリンク付け候補の単語を増やすアルゴリズム 1 により，事業リソースから因果リソースへのリンク付けを行った．

まず自治体事業名を形態素解析し名詞を抽出する．その際汎用的な後をストップワードとして削除している．次に各名詞の同義語を日本語 WordNet から取得し，リンク付けの候補を増やす．さらに，抽出した名詞の意義素を日本語 WordNet から取得する．意義素とは単語の意味を定義したものであり，その単語を使用した例文が含まれている．本研究では単語の意味を定義した文章のみを使用し，形態素解析により名詞を抽出する．このようにしてリンク付けの候補と成る単語を多く抽出し，これらの単語と同じ名称の都市問題およびその因果リソースのクラスと `dct:subject` プロパティでリンク付けを行う．

[8] https://data.city.osaka.lg.jp/
[9] https://poi.apache.org/

Algorithm 1 自治体事業リソースと都市問題因果関係リソースとのリンク付けアルゴリズム

Require: $businessList, causalityList, stopwordList$

Ensure: RDF

```
for each business ∈ businessList do
  morphemeList_business ⇐ morphologicalAnalysis(business.label)
  for each morpheme_business ∈ morphemeList_business do
    if morpheme_business.partOfSpeech == Noun
    && !stopwordList.contains(morpheme_business.text) then
      if causalityList.contains(morpheme_business.text) then
        causality ⇐ causalityList.get(morpheme_business.text)
        RDF.addStatement(business, dct : subject, causality)
      end if
      //Obtain synonym words of the noun words in the business name
      synonymList ⇐ getSynonymList(morpheme_business.text)
      for each synonym ∈ synonymList do
        if causalityList.contains(synonym) then
          causality ⇐ casaulityList.get(synonym)
          RDF.addStatement(business, dct : subject, causality)
        end if
      end for
      //Obtain short sentences describing the word sense
      gloss ⇐ getGloss(morpheme.text)
      morphemeList_gloss ⇐ morphologicalAnalysis(gloss)
      for each morpheme_gloss ∈ morphemeList_gloss do
        if morpheme_gloss.partOfSpeech == Noun
        && !stopwordList.contains(morpheme_gloss.text) then
          if causalityList.contains(morpheme_gloss.text) then
            causality ⇐ casaulityList.get(morpheme_gloss.text)
            RDF.addStatement(business, dct : subject, causality)
          end if
        end if
      end for
    end if
  end for
end for
```

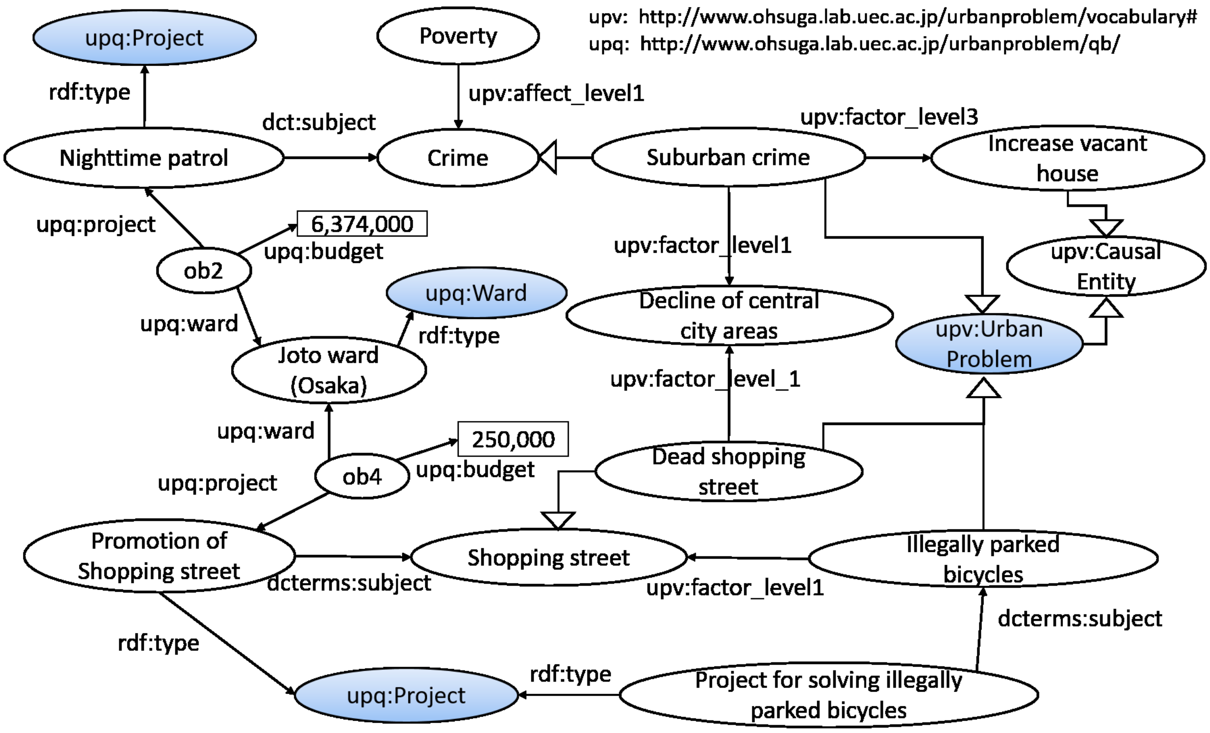

図 5.4: 最終的に構築される都市問題因果関係 LOD の一部

図 5.4 は最終的に構築される Linked Data（都市問題因果関係 LOD）の一部分である．トリプル数は 70,076 となった．構築した都市問題因果関係 LOD は我々のウェブサイト[10]で，Linked Data に対して検索が実行可能な SPARQL エンドポイントとともに公開している．

5.1.6 都市問題の因果関係抽出結果の考察

表 5.1 に最終的に使用した記事数，抽出文数，抽出単語数の合計を示す．5.1.2 節で述べたように要因に関しては “素因”，“導因” などの検索ノイズとなる類語を多く除き，影響の方が使用した類語が多かったため，文書，文，単語の数が多くなっている．また，クラウドソーシングの一致度を計るため，Fleiss の Kappa 係数 [20] を計算した．Kappa 係数は複数の項目における選択の一致率を算出できる指標であり，数値が高いほど選択の一致率が高いことを示す．本研究ではワーカ数は 1 ワードクラウドに付き 50 人であり，要因に関する一致度の平均は 0.291，影響に関する一致度は 0.212，全体の平均は 0.256 となった．Landis

[10] http://www.ohsuga.lab.uec.ac.jp/urbanproblem/

らによる Kappa 係数の解釈 [37] を表 5.2 に示す．この表によると，Kappa 係数 0.21 から 0.40 は “Fair” である．したがって，クラウドソーシングによるワードクラウドからの因果単語選択には弱から中程度の一致度が確認でき，本手法では弱から中程度ワーカーの共通認識がある因果単語が選択されたといえる．

一致度が弱から中程度にとどまった原因として，都市問題と関連しない単語が中心に集まり，本来関係のある単語が目立たなくなり選択されづらくなったことが考えられる．これは，行政の文書は冗長的な表現が多く複雑な重文が多く含まれていたため，係り受け解析を用いた因果単語候補の抽出が上手くいかない例があったことにより生じている．また，ワーカの予備知識が共通していないことも考えられ，都市問題の要因と影響の予備知識・問題意識が不足しているからこそ，都市問題が解決していないことも読み取れる．一方で，交通事故の要因に関しては 0.443 と比較的高い一致度が確認できた．これは，警視庁，教育機関，報道機関などが一般に向けて交通事故に関して説明する機会が多いため，ワーカの予備知識が共通していたのではないかと考える．

ただし，本手法でワーカによる単語選択に高い一致度が見られるということが，課題横断的なナレッジグラフの構築において必ずしも良い結果をもたらすとは言えず議論の余地がある．ワーカの単語選択の一致度が高いということは，対象の都市問題の因果関係として自明な事物ばかりが抽出されている可能性が考えられる．その結果を元にナレッジグラフを構築した際に，検索結果から自明な因果関係ばかりが得られることとなり，自治体及び都市問題の専門家にとっては意外性のある有益な情報を得られない可能性がある．一方，ワーカによる単語選択の一致度が著しく低いと，クラウドソーシングのタスクとして信頼性が低いことが言及され，その結果をもとに構築されるナレッジグラフの信頼性も低いと言える．したがって，本手法は単語選択の一致度に改善の余地を残しているが，意外性のある因果関係を抽出できるという観点ではある程度適切であったと言える．今後は自然言語からの因果単語抽出手法の改善，ワードクラウド以外の単語可視化方法，ワーカの段階的育成 [72] などを取り入れた上で，単語選択の一致度の高低がナレッジグラフの活用結果にどのように作用するか検証したい．

クラウドソーシングのタスクとして，要因と影響それぞれ 10 個ずつ選択してもらうタスクを設定したが，最後に自由記述欄を設けた．自由記述欄の回答の内訳を表 5.3 に示す．

表 5.1: 因果単語抽出結果

	都市問題の因果関係の記述を含む文書数	“要因” や “影響” の同義語を含む文数	抽出された単語数
要因	1,438	4,481	3,110
影響	2,465	9,082	4,661

表 5.2: Landis らによる Kappa 係数の解釈

Kappa Statistic	Strength of Agreement
<0.00	Poor
0.00–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost Perfect

最も多かった回答は “記述なし” であったが，次に多かった回答は対象の都市問題に関する自身の意見であり，これは自由記述欄に記述があった回答のうちの 62%にあたる．因果単語を選択した理由，選択したワード以外に候補として考えられるもの，解決に向けてどのような取組が必要であるかなどを記述するワーカがいた．これらは有用なコメントであり，今後ナレッジグラフに追加することを検討している．

5.2　都市問題の課題横断的なナレッジグラフの拡充と分析

本節では，構築したナレッジグラフの因果的な欠損補完による拡充と分析例について説明する．

5.2.1　推論ルールを用いた欠損因果リンクの補完

構築した都市問題因果関係 LOD には `UrbanProblem` クラス間や `CausalEntity` クラス間，および `UrbanProblem` と `CausalEntity` 間の因果関係が含まれているが，Web 情報か

表 5.3: 自由記述欄の回答の内訳

コメント内容	コメント数	比率
都市問題の因果関係や解決策などに関する意見や事例	82	9.6%
都市問題に関する感想	16	1.8%
クラウドソーシングタスクに関する感想	33	3.8%
特になし	719	84.5%

ら一文単位で半自動的に構築しているため，異なる文脈で出現する因果単語間の関係に欠損が生じるという問題がある．そこで，独自に定義した推論ルールにより欠損因果リンクを補完する手法を提案する．図 5.5 に上位クラス，代替リソースに基づいて補完される欠損リンクを示す．図 5.5 中の数字は次の SWRL ルールに対応している．薄字で示した番号は，その番号に対応する SWRL ルールと同等に補完されるものである．

```
(1) (?x upv:affect ?y), (?y skos:broader ?yp) -> (?x upv:
    probablyAffect ?yp) .
(2) (?x upv:affect ?y), (?x skos:broader ?xp) -> (?xp upv:
    probablyAffect ?y) .
(3) (?x upv:affect ?y), (?x skos:broader ?xp), (?y skos:broader ?yp)
   -> (?xp upv:mayAffect ?yp) .
(4) (?x upv:affect ?y), (?y prov:alternateOf ?yalt) -> (?x upv:
    likelyAffect ?yalt) .
(5) (?x upv:affect ?y), (?y skos:broader ?yp), (?yp prov:alternateOf ?
    ypalt) -> (?x upv:mightAffect ?ypalt) .
(6) (?x upv:affect ?y), (?x skos:broader ?xp), (?y prov:alternateOf ?
    yalt) -> (?xp upv:mightAffect ?yalt) .
(7) (?x upv:affect ?y), (?x skos:broader ?xp), (?y skos:broader ?yp),
    (?yp prov:alternateOf ?ypalt) -> (?xp prov:possiblyAffect ?ypalt) .
(8) (?z upv:affect ?yalt), (?yalt prov:alternateOf ?y), (?y skos:
    broader ?yp) -> (?z upv:mightAffect ?yp) .
(9) (?z upv:affect ?yalt), (?yalt prov:alternateOf ?y), (?y skos:
    broader ?yp), (?yp prov:alternateOf ?ypalt) -> (?z upv:
    possiblyAffect ?ypalt) .
(10) (?z upv:affect ?yalt), (?z skos:broader ?zp), (?yalt prov:
    alternateOf ?y), (?y skos:broader ?yp) -> (?zp upv:possiblyAffect ?
    yp) .
(11) (?z upv:affect ?yalt), (?z skos:broader ?zp), (?yalt prov:
    alternateOf ?y), (?y skos:broader ?yp), (?yp prov:alternateOf ?
    ypalt) -> (?zp upv:possiblyAffect ?ypalt) .
```

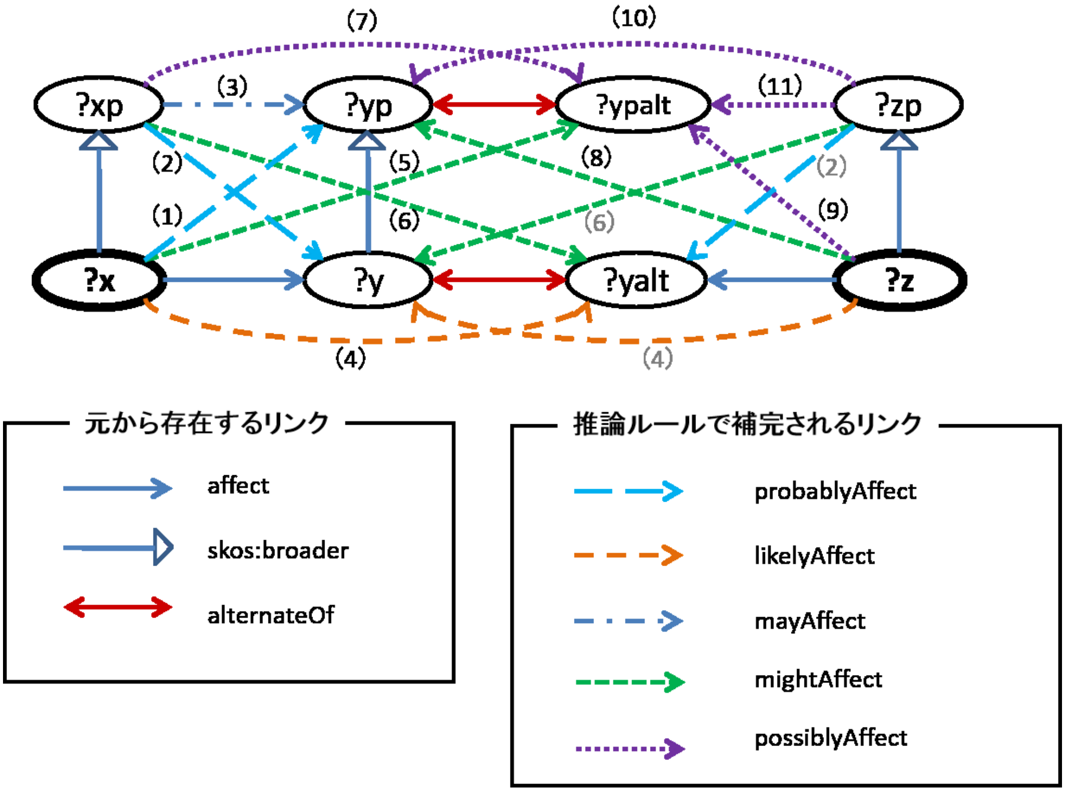

図 5.5: 推論ルールにより補完される欠損因果リンク

各ルールの意味は次の通りである．

(1)　「x が y に影響を与える」かつ「y の上位概念に yp がある」ならば「x は yp にも影響を与えるかもしれない」

(2)　「x が y に影響を与える」かつ「x の上位概念に xp がある」ならば「xp は y にも影響を与えるかもしれない」

(3)　「x が y に影響を与える」かつ「x の上位概念に xp がある」かつ「y の上位概念に yp がある」ならば「xp は yp にも影響を与えるかもしれない」

(4)　「x が y に影響を与える」かつ「y の代替関係に yalt がある」ならば「x は yalt にも影響を与えるかもしれない」

(5) 「x が y に影響を与える」かつ「y の上位概念に yp がある」かつ「yp の代替関係に ypalt がある」ならば「x は ypalt にも影響を与えるかもしれない」

(6) 「x が y に影響を与える」かつ「x の上位概念に xp がある」かつ「y の代替関係に yalt がある」ならば「xp は yalt にも影響を与えるかもしれない」

(7) 「x が y に影響を与える」かつ「x の上位概念に xp がある」かつ「y の上位概念に yp がある」かつ「yp の代替関係に ypalt がある」ならば「xp は ypalt にも影響を与えるかもしれない」

(8) 「z が yalt に影響を与える」かつ「yalt の代替関係に y がある」かつ「y の上位概念に yp がある」ならば「z は yp にも影響を与えるかもしれない」

(9) 「z が yalt に影響を与える」かつ「yalt の代替関係に y がある」かつ「y の上位概念に yp がある」かつ「yp の代替関係に ypalt がある」ならば「z は ypalt にも影響を与えるかもしれない」

(10) 「z が yalt に影響を与える」かつ「z の上位概念に zp がある」かつ「yalt の代替関係に y がある」かつ「y の上位概念に yp がある」ならば「zp は yp にも影響を与えるかもしれない」

(11) 「z が yalt に影響を与える」かつ「z の上位概念に zp がある」かつ「yalt の代替関係に y がある」かつ「y の上位概念に yp がある」かつ「yp の代替関係に ypalt がある」ならば「zp は ypalt にも影響を与えるかもしれない」

これらのルールにより，`probablyAffect`, `likelyAffect`, `mayAffect`, `mightAffect`, `possiblyAffect` の 5 つの因果プロパティが推論される．これらのプロパティの因果的な意味合いの強さは `probablyAffect` > `likelyAffect` > `mayAffect` > `mightAffect` > `possiblyAffect` としている．我々は `upv:affect` プロパティと `prov:alternateOf` プロパティのコストを 1，`skos:broader` および `rdfs:subClassOf` プロパティのコストを 0.75 として設定し，ルールにおける前件部分のプロパティのコストの合計に基づいて，後件のプロパティを決定した．定義したルールを RDF ストアの Stardog[11] に格納している．Stardog は推論エンジンの

[11] `https://www.stardog.com`

Pellet [56] を内蔵しており，OWL によるオントロジー推論と SWRL による独自定義のルールによる推論をサポートしている．しかし，可読性や記述の容易さ等の面から SPARQL 書式に基づく Stardog Rules Syntax (SRS) が Stardog で推奨されており，実際には我々は SWRL ルールを SRS に変換して格納している．

結果として，`probablyAffext` プロパティ1,058 個，`likelyAffect` プロパティ122 個，`mayAffect` プロパティ191 個，`mightAffect` プロパティ333 個，`possiblyAffect` プロパティ171 個が推論された．上述の通り，推論されたプロパティのうち `probablyAffect` が最も起こりうる可能性が高く，`possiblyAffect` が最も起こりうる可能性が低いことを意味しているが，これらの正しさについては長期的なデータ収集と統計的分析による評価が必要である．本手法では，起こりうる可能性として考えられる因果関係の候補を，ある程度説明可能な根拠（ルール）に基づいて抽出することに重きを置いている．したがって，都市問題の因果関係の再現率を高める手法であり，適合率を高める手法ではないことに注意されたい．

このように因果関係，階層関係，代替関係に基づいて新たな因果関係に関するトリプルを補完することにより，ナレッジグラフ内のある一つのエンティティを基準とした時に，間接的に因果関係のあるエンティティについて，中間エンティティへの経路へを細かく把握せずとも検索することが可能になる．例えば，放置自転車の影響として考えられる他のエンティティについては，次の SPARQL クエリで網羅的に検索することが可能である．

```
PREFIX upr: <http://www.ohsuga.lab.uec.ac.jp/urbanproblem/resource/>
PREFIX upv: <http://www.ohsuga.lab.uec.ac.jp/urbanproblem/vocabulary#>
SELECT * WHERE {
upr:Illegally_parked_bicycles (upv:affect|upv:probablyAffect|upv:
    likelyAffect|upv:mayAffect|upv:mightAffect|upv:possiblyAffect) ?o .
}
```

この SPARQL クエリにより，放置自転車の影響として直接的につながっているエンティティだけでなく，推論されたプロパティでつながっているエンティティもまとめて検索することができる．このように，推論ルールでエンティティ間の因果関係のバイパスが補完されることにより，より柔軟な検索が可能なナレッジグラフとなる．

5.2.2 都市問題解決施策の検討に向けたSPARQL検索

本節では構築したした都市問題因果関係LODを使用した検索例を提示し考察する．

図5.6はホームレスの因果関係とその予算情報について検索するクエリにより得られた部分グラフの例である．この図から，経済的困窮，雇用不足の他，ギャンブル依存等が原因でホームレスが増えていることがわかる．また，ホームレスは貧困ビジネスの発生を招き，貧困ビジネスは生活保護受給者の増加や地域差の増大や格差社会の助長につながることがわかる．貧困ビジネスとは経済的に困窮している貧困層を対象として利益を上げるビジネスであり，違法行為や非人道的な方法で金銭を巻き上げることが社会問題となっている．図の因果関係の具体例として，生活保護未受給の路上生活者に対して，生活保護の受給を強要してだまし取る貧困ビジネスが存在する．

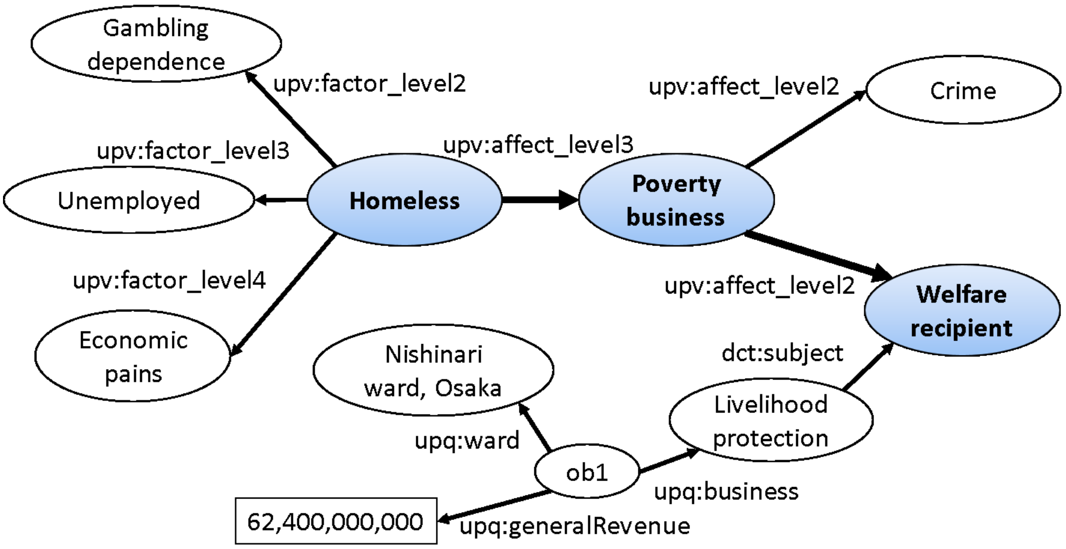

図5.6: ホームレスの因果関係の部分グラフ

2015年の統計データによると大阪市は生活保護費が2,907億円を超えており，国内で最も生活保護費が多い自治体である．中でも西成区の生活保護受給費は624億円を超えており，二位以下に2.5倍以上の差をつけている．西成区ではホームレスを対象とした貧困ビジネスの発生も問題視されており，アパートを契約させて生活保護費からピンハネする事例や，携帯の名義貸しと引き換えに報酬を受け取り，後で高額な請求が来る事例が問題となっている．このようなホームレスから始まる負の連鎖を止めるために，ホームレスの原

因となっている雇用不足やギャンブル依存の対策を各自治体で進めることが望ましい．しかし，現在西成区だけでなく大阪市全体でもこのような取り組みは行われていない．雇用不足やギャンブル依存の対策事業を行うことで，生活保護受給者の減少および生活保護費の削減につながると考えられる．

図5.7はポイ捨ての原因と影響およびそれらに関連する予算情報を検索するクエリにより得られた部分グラフである．この図から，灰皿設置（不足）がポイ捨てに影響を与えており，ポイ捨ては地下水汚染を引き起こすことがわかる．また，ある事物により生じる影響は対象先の事物の上位概念に伝搬するという5.2.1節のルール(1)により，ポイ捨てから地下水へ`affect`プロパティによる直接的な関係が推論できる．さらに，地下水は騒音に影響を与えることが，騒音に関する因果関係抽出結果から得られている．結果として，灰皿設置（不足）はポイ捨てを引き起こすだけでなく，連鎖的に地下水汚染および騒音に影響を与える可能性があることが`affect`プロパティの推移律により判明した．この因果関係の具体例としては，路上喫煙者がポイ捨てしたタバコが地下水に流れ込むことで地下水が汚染されたり，地下水の水質汚濁改善のため地下で配管工事が行われ，騒音被害が発生することなどが挙げられる．

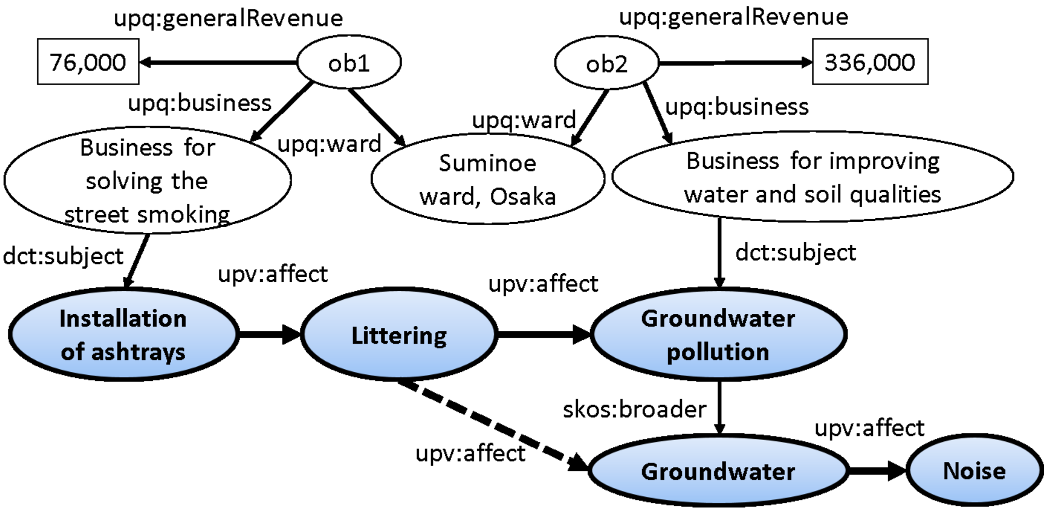

図5.7: ポイ捨ての因果関係の部分グラフ

都市問題とその予算の関係に関しては，大阪市住之江区は土壌汚染・水質汚濁対策事業に区・シティマネージャー自由経費予算として336,000円を投じていることがわかる．一

方で同区において路上喫煙対策事業は 76,000 円しか予算を投じていないことがわかる．そこで，一連の都市問題のより上流に位置する路上喫煙対策事業予算を増額することでより効果的な問題解決に繋がる可能性がある．路上喫煙対策事業の予算をより多く投じることで，路上喫煙の減少だけでなく，連鎖的に引き起こされるポイ捨て，地下水汚染，騒音問題の改善が期待される．一方で，地下水汚染が軽減されれば，土壌汚染・水質汚濁改善事業は予算を削減できる可能性がある．

5.2.3 都市問題の因果関係からなる悪循環発見の試み

本節では，SPARQL を用いて都市問題因果関係 LOD から都市問題の因果関係の悪循環を発見する手法について説明する．本研究で発見する悪循環は，`affect` プロパティのサブプロパティおよび前節で推論されたプロパティからなるサイクルであり，少なくとも一つの都市問題リソース（都市問題クラスのサブクラス）を含む．サイクルのサイズは 3 ノードから 6 ノードまでとしている．下記に都市問題の悪循環発見のための SPARQL クエリを示す．

```
PREFIX upv: <http://www.ohsuga.lab.uec.ac.jp/urbanproblem/vocabulary#>
SELECT DISTINCT ?x ?y ?z WHERE {
?x        upv:affect        ?y .
?y        upv:affect        ?z .
?z        upv:affect        ?x .
?x        rdfs:subClassOf   ?upv:UrbanProblem .
FILTER(?x != upv:UrbanProblem && ?y != upv:UrbanProblem && ?z != upv:
    UrbanProblem)
FILTER(?x != upv:CausalEntity && ?y != upv:CausalEntity && ?z != upv:
    CausalEntity)
FILTER(?x != ?y && ?x != ?z && ?y != ?z)
}
```

推論ルール適用してこのクエリを実行することで悪循環を発見する．また，発見した悪循環には `Poverty` $\xrightarrow{\text{affect}}$ `Truancy` $\xrightarrow{\text{affect}}$ `Disease` と，`Truancy` $\xrightarrow{\text{affect}}$ `Disease` $\xrightarrow{\text{affect}}$ `Poverty` のような重複が含まれるため，これらの重複を削除する．

表 5.4 は発見された悪循環の数である．結果として，推論ルールによる欠損リンクの補完前では 951 の悪循環が発見され，欠損リンクの補完後では 1,904 の悪循環が出された．

表 5.4: 発見された悪循環の数

悪循環	推論ルール無し	推論ルール有り
3ノード	33	45
4ノード	168	308
5ノード	236	460
6ノード	514	1,091
合計	951	1,904

5.2.4　悪循環につながる根源的課題の発見

都市問題の悪循環の発生を阻止するためには，その悪循環の発生につながる根源的課題を把握し対策を講じることが重要である．特に複数の悪循環の発生につながる事物は対策が急がれる課題である．そこで，この根源的課題のグラフパターンを定義し，SPARQLにより発見する実験を行った．本研究では図5.8のように根源的課題として２つのパターンを定義する．図5.8における左図は複数の悪循環に影響を与えるタイプの根源的課題であり，次のSPARQLクエリにより発見する．

```
PREFIX upv: <http://www.ohsuga.lab.uec.ac.jp/urbanproblem/vocabulary#>
SELECT DISTINCT ?x ?y ?z ?a ?b ?c WHERE {
?x      upv:affect      ?y .
?y      upv:affect      ?z .
?z      upv:affect      ?x .
?a      upv:affect      ?b .
?b      upv:affect      ?c .
?c      upv:affect      ?a .
?root   upv:affect      ?x ;    upv:affect      ?a .
?y      rdfs:subClassOf ?yp .
?b      rdfs:subClassOf ?bp .
filter(?yp = upv:UrbanProblem)
filter(?x != ?y && ?x != ?z && ?y != ?z)
filter(?bp = upv:UrbanProblem)
filter(?x != ?b && ?x != ?c && ?b != ?c && ?y != ?b && ?z != ?c)
}
```

図5.8の右図は複数の悪循環内に共通するタイプの根源的課題であり，次のSPARQLクエリにより発見する．

```
SELECT DISTINCT ?x ?y ?z ?b ?c WHERE {
?x        upv:affect       ?y .
?y        upv:affect       ?z .
?z        upv:affect       ?x .
?x        upv:affect       ?b .
?b        upv:affect       ?c .
?c        upv:affect       ?x .
?y        rdfs:subClassOf ?yp .
?b        rdfs:subClassOf ?bp .
filter(?yp = upv:UrbanProblem)
filter(?x != ?y && ?x != ?z && ?y != ?z)
filter(?bp = upv:UrbanProblem)
filter(?x != ?b && ?x != ?c && ?b != ?c && ?y != ?b && ?z != ?c)
}
```

悪循環のサイズは 3 ノードから 6 ノードまで変化する．結果として合計 144 パターンがマッチし，根源的課題は 28 種類発見された．

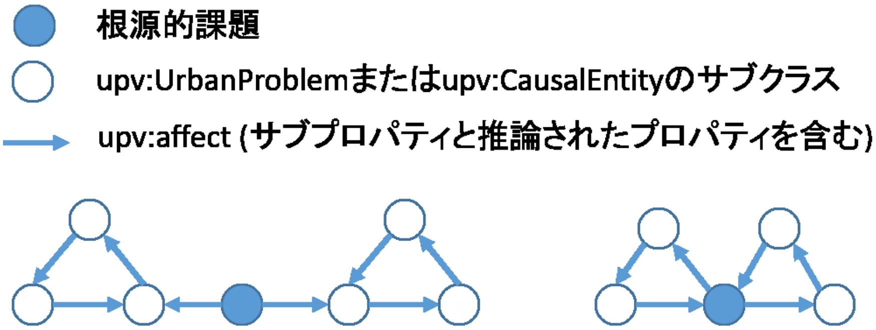

図 5.8: 根源的課題のグラフパターン

5.2.5 評価

本研究では，悪循環を構成する因果関係トリプルに誤りがない場合，その悪循環が起こりうる可能性があるとした．我々の知る限り都市問題に関する因果関係の正解データは存在しないため，都市問題の専門家の方々に協力を頂き，まず因果関係トリプルの評価を行った．全ての都市問題の悪循環を評価することは困難であるため，本研究ではホームレス問題と犯罪に関する悪循環を評価対象とし，ホームレス支援に取り組む NPO 法人，再犯問

題の解決に取り組む企業，都市問題の調査研究を行っている大阪市政調査会から計 4 人の専門家を集め，大阪市市民局の協力のもとでディスカッション形式で評価を行った．図 5.9 は評価の際の様子である．専門家の方々には，各因果関係について，(1) 正しいといえる，(2) 起こりうる可能性がある，(3) 誤っている，の 3 つの内のどれに該当するか議論していただいた．さらに，本研究で抽出した因果関係以外にも追加すべき因果関係がある場合は，その内容についてコメントを頂いた．結果として，ホームレスと犯罪に関する因果関係について (1) は 21 個，(2) は 154 個，(3) は 5 個，追加すべき因果関係について 14 個の意見を頂いた．例えば，(1) については多重債務がホームレスの一因となっていることは確かであるため，多重債務 $\xrightarrow{\text{affect}}$ ホームレス等が該当した．(2) については，ホームレスの人々が廃品回収時にプラスチックゴミを散らかして環境汚染につながる可能性があるという理由から，ホームレス $\xrightarrow{\text{affect}}$ 環境汚染等が該当した．(3) については，ホームレスの高齢化は問題になっているが，高齢化がホームレスの原因になっているわけではないとの理由から，高齢化 $\xrightarrow{\text{affect}}$ ホームレス等が該当した．追加すべき因果関係については，核家族化が犯罪の原因の一つになっているなどの意見を頂いた．

次に，発見した悪循環について評価した．本研究では，ホームレス問題と犯罪に関する因果関係について専門家から頂いた (1) または (2) の意見に該当するもののみから構成される悪循環を，起こりうる可能性がある悪循環 (正解悪循環) とした．結果としてホームレスと犯罪に関する 196 個の正解悪循環が発見された．例えば，貧困 $\xrightarrow{\text{affect}}$ 貧困ビジネス $\xrightarrow{\text{affect}}$ 日雇い労働 $\xrightarrow{\text{affect}}$ ホームレス $\xrightarrow{\text{affect}}$ 病気などが挙げられる．しかしながら，この悪循環が実際に観測されるかどうかは長期的な調査が必要である．

5.2.6　考察

5.2.4 章の実験により発見された根源的課題は 28 種類の内，正解悪循環に影響を与えるものとして例えば不就学問題があった．不就学問題は，貧困 $\xrightarrow{\text{affect}}$ ホームレス $\xrightarrow{\text{affect}}$ 貧困ビジネス $\xrightarrow{\text{affect}}$ 日雇い労働の正解悪循環と $\xrightarrow{\text{affect}}$ 治安悪化 $\xrightarrow{\text{affect}}$ 落書き $\xrightarrow{\text{affect}}$ 窃盗の悪循環に影響を与える可能性がある．想定されるシナリオとして，不就学児童は安定した職業に就くことができず，将来的に貧困になる可能性があり，貧困の悪循環に陥る可能性もある．また，不就学児童の非行により治安が悪化する可能性があり，治安悪化の悪循環が発

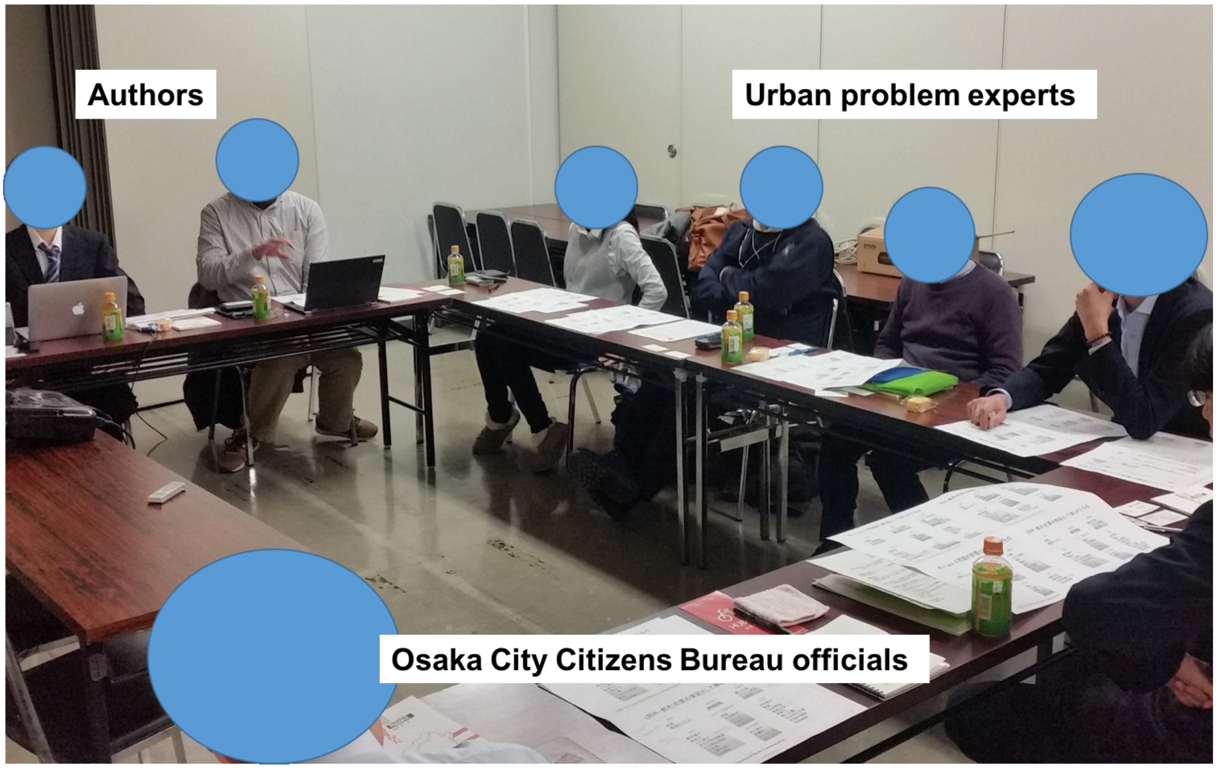

図 5.9: 専門家による都市問題の因果関係評価の様子

生する可能性がある．これらの事象が実世界で観測されるかどうかは長期的な調査が必要であるが，解決が急がれる課題の抽出および検討の際に利用できると考える．

最後に，都市問題の専門家および大阪市市民局から，本研究は都市問題の全体像を把握するのに有用であるとのコメントを頂いた．専門家は通常一つの都市問題に注目して課題解決を考えるため，本研究により他の問題との因果関係を見ることで視野が広まったとのコメントを頂いた．また，構築したナレッジグラフは課題解決に向けた専門家同士の議論の際のツールとして有用であるとのコメントを頂いた．本研究で構築したナレッジグラフを議論の際のツールとして使用する際には，都市問題間の因果関係について，その導出根拠を元データとともに提示し説明する機能が重要である．課題横断的なナレッジグラフは文書を元に構築しているため，因果関係導出の根拠となる文を提示することが可能である．さらに，因果関係抽出プロセスにおける，ワードクラウドから単語選択を行うクラウドソーシングの結果について，選択したワーカ数や選択の一致率などを同時に提示することが可能である．さらに，クラウドソーシングの際に自由記述欄から得られた回答には，なぜそ

の単語を選択したか理由付きで述べているものも存在するため，合わせて提示することが可能である．一方で，都市問題に関する統計情報に関しては，元となる文書に自然文として含まれていなければ，現状で抽出はできていない．説明材料として，文だけでなく関連する統計情報も提示する機能を実現するためには，今後データソースを拡張していく必要がある．都市問題解決に関わる様々な立場の人は各々異なる見方をすることが予想されるため，因果関係の結果だけを切り取って提示するのではなく，このような説明機能を実装することが課題である．

また，専門家からは因果関係について数値的な情報もあると便利であるというコメントも頂いた．数値情報の付加については，確率的な値である場合はより多くのデータ収集が必要であるため，今後の課題としたい．推論により補完される因果関係の曖昧性の度合いについては，Fuzzy Ontology [7] 等の適用を検討している．

本章では課題横断的なナレッジグラフを構築したが，本章冒頭で述べたように都市問題ごとのナレッジグラフとクラスレベルで結合しているため，因果関係プロパティおよびクラス-インスタンス関係をたどる検索を行い，都市問題ごとのナレッジグラフから得られる詳細情報をマッシュアップして分析することが可能である．都市問題ごとのナレッジグラフについて，放置自転車の場合は対象とする問題をイベントと捉えて Event Ontology に基づくスキーマ構造となっている．したがって，`event:factor` プロパティをたどることで，課題横断的なナレッジグラフよりも時空間的に詳細かつ短期的な視点で，個別の要因となりうる情報を把握することが可能である．対して，課題横断的なナレッジグラフで得られる要因は都市問題間の関係であり，より抽象的かつ長期的な視点の情報が多い．本研究の枠組みを横展開してナレッジグラフを拡充していくことで，課題横断的なナレッジグラフの抽象的な因果関係をたどり，得られた都市問題のナレッジグラフのペアを組み合わせて分析することで，因果関係の根拠を理解することができると考えられる．また，その結果を受けて課題横断的なナレッジグラフの因果関係を修正するなど，継続的な知識洗練を行うことが課題である．

第6章　結論

本章では，本研究における課題，提案手法についてまとめ，最後に今後の課題について記述する．

6.1　まとめ

本研究では，データ利活用による都市問題の解決に貢献すべく，都市問題のデータ分析活用基盤としてナレッジグラフを設計・構築・拡充する手法を開発し，分析例を示した．本研究では，特に (1) 都市問題ごとに時間的・空間的に詳細なナレッジグラフと，(2) 複数の都市問題間の複雑な関係を記すナレッジグラフの 2 種類のナレッジグラフを構築している．

(1) に関しては，個別の都市問題の一例として放置自転車問題に焦点を当て，時間的・空間的に詳細な分析を可能にするためのスキーマを設計し，ソーシャルセンサからの実データ収集によりナレッジグラフを LOD として構築・公開した．しかし，このナレッジグラフにはある時刻における観測値の欠損（時間的欠損値）とある場所における観測値の欠損（空間的欠損値）の 2 種類の欠損が生じている．そこで，時間的欠損値については，ナレッジグラフから属性選択をしたベクトルデータを学習データとしてベイジアンネットワークを適用することで，約 70%の精度で放置台数を推定し補完した．また，空間的欠損値については，人の流れを流体に例えて都市の気流シミュレーションを行い，気流のよどみ点検出と周辺 POI のセマンティクス利用による推定手法を提案した．これにより，気流シミュレーションを用いた推定地点と観測値に相関があることが明らかになり，ランダム推定より有意に高い精度で空間的欠損値（観測地点）を推定し補完した．

さらに，他の都市問題としてゴミのポイ捨て問題を対象として LOD を設計・構築し，空間的欠損値（未観測地域をグリッド分割した際の 1 セルにおけるポイ捨てゴミの個数）を推定する手法を提案した．これは上述の空間的欠損値推定手法を格子単位に行うように発

展させたものである．気流シミュレーション結果と地理空間情報を含むナレッジグラフを構築し，RDF2vecによりナレッジグラフを500次元のベクトル空間に埋め込むことで，このベクトルデータを元にニューラルネットワークを用いてポイ捨てゴミの個数を予測するグラフベース手法を開発した．また，気流シミュレーション結果を地図上にヒートマップとして可視化した画像を元に，CNNによりゴミの個数を予測する画像ベース手法も開発した．結果としてグラフベース手法のRMSE値1.5，画像ベースのRMSE値1.47と，ほぼ同じ程度に高精度で予測することが可能になったことを確認した．また，グラフベース手法と画像ベース手法をニューラルネットワークモデル上で結合する手法を提案し，RMSE値0.24と高精度で予測することが可能になった．ポイ捨てや放置自転車のように，都市空間と人の行動に関係の深い問題に関して，本研究で提案したナレッジグラフの時間的・空間的な欠損推定手法を同様に適用できる可能性がある．

構築したナレッジグラフの分析例として，SPARQLを用いた検索による分析結果を示した．さらに，ナレッジグラフを活用して放置自転車の時間的変化を可視化するアプリケーションを開発し，東京都青少年・治安対策本部に協力する形で課題解決促進に貢献した．

(2)に関しては，都市問題の因果関係から自治体予算事業までの検索を可能にするスキーマを設計した．次に，複数の都市問題に関する因果関係として考えられる単語を，Web上に存在する記事，SNS情報，自治体オープンデータ，学術文献等から自然言語処理技術を用いて抽出した．さらに，ワードクラウドとクラウドソーシングを用いた因果単語の絞り込みを行い，この単語を元にナレッジグラフをLODとして構築・公開した．自治体予算事業に関しては大阪市を例として，表データから設計したスキーマに基づいてLOD化し，都市問題の因果関係リソースと結合した．さらに，異なる文脈で抽出される単語の因果関係の欠損に着目し，11パターンのSWRLルールを用いた推論により因果的な欠損を補完した．

構築したナレッジグラフの分析例としてSPARQLを用いた検索結果を示し，都市問題間の因果関係と自治体予算情報から考えられる解決施策の例について考察した．また，SPARQLを用いて都市問題の因果関係からなる悪循環を発見する実験を行い，1904個の悪循環を発見した．この結果を評価するために，大阪市市民局協力のもと，都市問題の中でもホームレス問題と犯罪に関する専門家による因果関係の評価を行った．結果として，ホームレス

と犯罪に関する 180 個の因果関係のうち，正しい因果関係 21 個と，起こりうる可能性のある因果関係 154 個が判明した．また，これらの因果関係からなる起こりうる可能性のある悪循環を 196 個発見し，その例を考察した．構築したナレッジグラフに関して，都市問題の専門家および大阪市市民局から，都市問題の全体像を把握する際や，専門家同士の議論活性化のツールとして有用であるとのコメントを頂き，都市問題解決の一助として本研究の有用性が確認できた．

このように，本研究では，都市問題自体の詳細なデータが不足しているという課題に対して，(1) のナレッジグラフの設計・構築・欠損補完拡充・分析活用の一連の手法を提案し，都市問題という大きな問題の中でも，特に地理空間的特徴と人の行動に関連の深い都市問題という範囲で横展開できる．さらに，都市問題の因果が複雑に絡み合う構造の整理と分析という課題に対して，(2) のナレッジグラフの設計・構築・欠損補完拡充・分析活用の一連の手法を提案し，これは対象範囲としてより広義の都市問題に一般化できる．本研究はこれらの範囲における枠組みを提案しており，データ利活用による都市問題解決の一助となることを目的とし，その有用性を確認できた．

6.2　今後の課題

本研究を通して得られた課題として，都市問題ごとのナレッジグラフの構築に関してはデータ収集のためのインセンティブ設計が挙げられる．本研究では，初めに報酬型クラウドソーシングによるデータ収集を試みたがうまく行かず，SNS を利用して課題意識のある人を主な対象として呼びかけることでデータを収集した．データ収集結果がすぐに LOD に追加されて可視化されるアプリケーションを開発し，東京都のサイト上からも紹介することでデータ収集への協力を仰いだ．しかし，さらに継続的にデータを収集するためには，ユーザへのインセンティブ付与の設計が必要である．例えば，ゲーミフィケーション要素を取り入れたアプリケーションを開発することや，定期的にオープンデータソンのようなデータ投稿型のイベントを開催するなどの解決策が考えられる．

また，都市問題ごとのナレッジグラフのスキーマ設計に関しては，放置自転車 LOD において factor プロパティの値の粒度が異なるという問題がある．本研究では，スキーマ設計における要件抽出の際に factor の値を手動でクラスタリングしているが，各リソースを

上位オントロジーに基づいて階層化することや，情報量などの定量的指標を用いて粒度をあわせるなどの工夫が必要であると考えている．

ナレッジグラフの時間的欠損値の推定に関しては，精度向上に向けていくつかの課題が得られた．まず，データ数が然程多くないためベイジアンネットワークにおける学習が充分でないことが挙げられる．これは先に述べたデータ収集の課題と関連し，継続的なデータ収集手法の設計によりある程度改善できると考えている．また，ソーシャルセンサにより得られたデータのため，均等に観測したデータではなく，観測地点や観測値によってデータの偏りがある．観測地点の偏りに関しては，東京都だけでなく他の自治体との連携により，データ収集を行っていくことである程度改善できると考えている．観測値の偏りに関しては，ソーシャルセンサゆえに観測者の行動時間に依存してしまう問題や，観測地点の範囲決定に個人差があったことも原因であると考えられる．そこで，より多くのユーザの獲得に加えて，観測地点に明確な範囲を定めるため，データ投稿時にアプリ上で範囲円を表示させるなどの修正を検討している．

空間的欠損値の推定に関しては，気流シミュレーションと周辺施設のセマンティクスを用いた推定値と，放置自転車の観測地点に相関があることが明らかになった．しかし，推定精度にはさらなる向上の余地が多く見られる．放置自転車の場合，駅に近づくほど自転車の放置地点が多く，駅から遠ざかるにつれて放置地点は減っていく．このような特性を考えて，駅中心からの距離に対応する確率的な重みを設定して予測するなどの改善策が考えられる．また，11 個のよどみ点もパターンを定義したが，よどみ点の大きさを 1 格子点としているため，細かな単位でのよどみ点検出となっている．実際には放置自転車は複数台が集まっていることから，よどみ点に一定の大きさを与えた上でのパターンを新たに定義し，実験する必要があると考えている．

また，ゴミのポイ捨て調査結果の LOD の空間的欠損値の推定に関して，Graph Embedding を用いたグラフベース手法では，CNN を用いた画像ベース手法に比べて数値情報をうまく学習できていない可能性がある．そこで，より細かく格子分割をすることで，1 セルあたりが持つ CFD シミュレーション結果の情報量を多くすることを検討している．しかしながら，格子分割を細かく行うことにより CFD シミュレーションの計算が発散してしまうため，最適なパラメータの調整が必要である．また，数値データに関して等間隔分類

の間隔調整や，自然分類や k-means などの数値分類手法の比較も課題である．また，他地域のゴミのポイ捨て状況を予測するには，データ数が十分でないという問題もある．本研究で行った実験では 3 駅分のデータを使用したが，今後対象地域数を増やしてデータを学習することで，未観測地点におけるデータ推定を高精度で行うことができると考える．ポイ捨て問題における時間的な欠損値については，時系列データではないことから放置自転車と同様の手法を適用することは困難であった．今後，時系列データを転移学習させることで，時間的な予測を行うことができるか検討する．

都市問題の課題横断的なナレッジグラフの構築に関しては，因果関係抽出に関して課題が残る．本研究では，因果関係に関する単語を含む文書から，該当文のみを対象として係り受け解析による因果単語候補を抽出し，クラウドソーシングを用いた絞り込みを行った．しかし，実際の文書には因果関係に関して文をまたいでの記述や，箇条書きによる記述，異なる段落での記述など様々な記述方法が存在する．したがって，これらの複雑な因果関係の記述の抽出に対応するために，CNN を用いた因果関係抽出 [38] のようなニューラル自然言語処理手法を適用を検討する必要がある．

1.2 節で述べたように，本研究における因果関係とは，単にある事物の要因や影響として指摘されうる関係であり，その他の科学的な意味合いを指していない．したがって，本研究で構築した課題横断的なナレッジグラフは，Web 上にある人間の知見の集合とも言えるデータ分析基盤であり，機械学習的アプローチの相補的な位置づけに当たる．次の課題として，各都市問題の因果関係の確からしさの算出が必要である．今後，都市問題ごとに長期的なデータ収集を行い，本研究のようなナレッジグラフの構築と時間的・空間的な拡充を行うことで，ナレッジグラフを活用した統計的分析を通じて各因果関係の確からしさを算出できると考えている．

さらに，今後，都市問題解決に向けた自治体や NPO 法人との連携や，市民協働イベントにおける研究成果の提供を通じて，本研究を本格的に社会実装していくことが課題である．

謝辞

本研究は JSPS 科研費 16K00419, 16K12411, 17H04705, 18H03229, 18H03340, 18K19835, 18J13988，および電気通信普及財団の助成を受けたものです．

本研究にあたり，ご多忙の中適切なご指導をくださった主任指導教員の川村隆浩客員准教授を初め，大須賀昭彦教授，田原康之准教授，清雄一准教授に感謝いたします．特に川村客員准教授には個別ゼミでの集中的なご指導に加えて，本務中や出張中においても大変お世話になり深謝申し上げます．共同研究者である大阪大学産業科学研究所の古崎晃司准教授には，研究における貴重なご意見とご指導に加えて，大阪出張の際に何度もお世話になったことを感謝申し上げます．また，学部時代の卒業論文指導から公私共にお世話になっている法政大学藤井昭博教授に感謝申し上げます．研究生活において，様々な面でお世話になりました大須賀・清・田原研究室の秘書の皆様，研究室の先輩，後輩，同期，OBOG の皆様，早稲田大学本位田真一教授をはじめ活発な議論と貴重な御意見を頂いた研究グループの皆様に感謝の意を表します．また，大阪市市民局の渡邊芳枝様を初め，評価実験にご協力いただきました市民局，ICT 戦略室の皆様，NPO 法人 Hoomedoor 代表の川口加奈様，NPO 法人釜ヶ崎支援機構の横田和彦様，株式会社ヒューマンハーバー大阪代表の黒川洋司様，大阪市政調査会事務局長の福田弘様に感謝申し上げます．また，放置自転車問題解決に向けて協力いただきました東京都総務局の前島肇様，東京都青少年・治安対策本部の畑仲智史様に感謝申し上げます．また，ゴミのポイ捨てデータを提供していただき，LOD 化実現にあたり密に議論をさせて頂きました株式会社ピリカ代表取締役の小嶌様，高橋様に感謝申し上げます．また，データ収集にご協力いただいたアプリケーションユーザの皆様，クラウドワーカの皆様に感謝申し上げます．さらに，LOD 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，NPO 法人リンクト・オープン・データ・イニシアティブの皆様に感謝申し上げます．そして，投稿した論文等に対して，国内外の多くの査読者から様々なコメントをいただき，大いに研究の参考になりました．研究コミュ

ニティの皆様に感謝申し上げます．審査を快く引き受けてくださいました大学院情報理工学研究科の田中健次教授，西野哲朗教授，柏原昭博教授に感謝申し上げます．最後に，これまであらゆる面で支えてくれた家族に感謝いたします．

参考文献

[1] Michael Ashburner and et.al. Gene ontology: tool for the unification of biology. *Nature Genetics*, Vol. 25, No. 1, pp. 25–29, 2000.

[2] 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.

[3] I. Augenstein, S. Padó, and S. Rudolph. Lodifier: Generating linked data from unstructured text. In *Proc. of the 9th Extended Semantic Web Conference (ESWC)*, pp. 210–224, 2012.

[4] 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.

[5] Robert Battle and Dave Kolas. Geosparql: enabling a geospatial semantic web. *Semantic Web Journal*, Vol. 3, No. 4, pp. 355–370, 2011.

[6] Stefan Bischof, Christoph Martin, Axel Polleres, and Patrik Schneider. Collecting, integrating, enriching and republishing open city data as linked data. In *International Semantic Web Conference*, pp. 57–75. Springer, 2015.

[7] Fernando Bobillo and Umberto Straccia. Fuzzy ontology representation using owl 2. *International Journal of Approximate Reasoning*, Vol. 52, No. 7, pp. 1073–1094, 2011.

[8] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pp. 1247–1250. AcM, 2008.

[9] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pp. 2787–2795. Curran Associates, Inc., 2013.

[10] Irene Celino, Dario Cerizza, Simone Contessa, Marta Corubolo, Daniele DellAglio, Emanuele Della Valle, and Stefano Fumeo. Urbanopoly–a social and location-based game with a purpose to crowdsource your urban data. In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, pp. 910–913. IEEE, 2012.

[11] Irene Celino, Simone Contessa, Marta Corubolo, DanieleDell ' Aglio, Emanuele Della Valle, Stefano Fumeo, Thorsten Krüger. Linking smart cities datasets with human computation–the case of urbanmatch. In *International Semantic Web Conference*, pp. 34–49. Springer, 2012.

[12] Michael Compton, Payam Barnaghi, Luis Bermudez, Raúl García-Castro, Oscar Corcho, Simon Cox, John Graybeal, Manfred Hauswirth, Cory Henson, Arthur Herzog, Vincent Huang, Krzysztof Janowicz, W. David Kelsey, Danh Le Phuoc, Laurent Lefort, Myriam Leggieri, Holger Neuhaus, Andriy Nikolov, Kevin Page, Alexandre Passant, Amit Sheth, and Kerry Taylor. The ssn ontology of the w3c semantic sensor network incubator group. *Web Semantics: Science, Services and Agents on the World Wide Web*, Vol. 17, pp. 25 – 32, 2012.

[13] Andrew Crapo and Abha Moitra. Toward a unified english-like representation of semantic models, data, and graph patterns for subject matter experts. *International Journal of Semantic Computing*, Vol. 7, No. 03, pp. 215–236, 2013.

[14] Gerben KD de Vries. A fast approximation of the weisfeiler-lehman graph kernel for rdf data. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 606–621. Springer, 2013.

[15] Gianluca Demartini, Djellel Eddine Difallah, and Philippe Cudré-Mauroux. Zencrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking. In *Proceedings of the 21st international conference on World Wide Web*, pp. 469–478. ACM, 2012.

[16] Dongpo Deng, Guan-Shuo Mai, Tyng–Ruey Chuang, Rob Lemmens, and Kwang-Tsao Shao. Social web meets sensor web: From user-generated content to linked crowdsourced observation data. In *Proceedings of the Workshop on Linked Data on the Web co-located with the 23rd International World Wide Web Conference (WWW 2014)*, pp. 1–10, 2014.

[17] Lian Duan, Tao Hu, En Cheng, Jianfeng Zhu, and Chao Gao. Deep convolutional neural networks for spatiotemporal crime prediction. In *2017 International Conference on Information and Knowledge Engineering (IKE)*, pp. 61–67, 2017.

[18] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, Vol. 12, No. Jul, pp. 2121–2159, 2011.

[19] Lorena Etcheverry, Alejandro Vaisman, EstebanZim￥'anyi. Modeling and querying data warehouses on the semantic web using qb4olap. In *Proc. of the 16th International Conference on Data Warehousing and Knowledge Discovery (DaWaK)*, pp. 45–56, 2014.

[20] J. L. Fleiss and J. Cohen. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, Vol. 33, No. 3, pp. 613–619, 1973.

[21] 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.

[22] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 855–864. ACM, 2016.

[23] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, Vol. 11, No. 1, pp. 10–18, 2009.

[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[25] María Herrero-Zazo, Isabel Segura-Bedmar, Janna Hastings, and Paloma Martínez. Dinto: using owl ontologies and swrl rules to infer drug–drug interactions and their mechanisms. *Journal of chemical information and modeling*, Vol. 55, No. 8, pp. 1698–1707, 2015.

[26] Minh X Hoang, Yu Zheng, and Ambuj K Singh. Fccf: forecasting citywide crowd flows based on big data. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, p. 6. ACM, 2016.

[27] Johannes Hoffart, Fabian M Suchanek, Klaus Berberich, Edwin Lewis-Kelham, Gerard De Melo, and Gerhard Weikum. Yago2: exploring and querying world knowledge in

time, space, context, and many languages. In *Proceedings of the 20th international conference companion on World wide web*, pp. 229–232. ACM, 2011.

[28] K. Höffner, M. Martin, and J. Lehmann. Linkedspending: Openspending becomes linked open data. *Semantic Web Journal*, Vol. 7, No. 1, pp. 95–104, 2016.

[29] Soyoung Hwang and Donghui Yu. Gps localization improvement of smartphones usingbuilt in sensors. *International Journal of Smart Home*, Vol. 6, No. 3, pp. 1–8, 2012.

[30] Krzysztof Janowicz, Armin Haller, Simon JD Cox, Danh Le Phuoc, and Maxime Lefrancois. Sosa: A lightweight ontology for sensors, observations, samples, and actuators. *Journal of Web Semantics*, 2018.

[31] G.F. Jenks. The data model concept in statistical mapping. *International yearbook of cartography*, Vol. 7, No. 1, pp. 186–190, 1967.

[32] Fumihiro Kato, Hideaki Takeda, Seiji Koide, and Ikki Ohmukai. Building dbpedia japanese and linked data cloud in japanese. In *2013 Linked Data in Practice Workshop (LDPW2013)*, pp. 1–11, 2014.

[33] Mayank Kejriwal and Pedro Szekely. Neural embeddings for populated geonames locations. In *International Semantic Web Conference*, pp. 139–146. Springer, 2017.

[34] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[35] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[36] Taku Kudo and Yuji Matsumoto. Japanese dependency analysis using cascaded chunking. In *proceedings of the 6th conference on Natural language learning-Volume 20*, pp. 1–7. Association for Computational Linguistics, 2002.

[37] J Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *biometrics*, pp. 159–174, 1977.

[38] Pengfei Li and Kezhi Mao. Knowledge-oriented convolutional neural network for causal relation extraction from natural language texts. *Expert Systems with Applications*, Vol. 115, pp. 512–523, 2019.

[39] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning entity and relation embeddings for knowledge graph completion. In *AAAI*, Vol. 15, pp. 2181–2187, 2015.

[40] Vanessa Lopez, Spyros Kotoulas, Marco Luca Sbodio, Martin Stephenson, Aris Gkoulalas-Divanis, and Pól Mac Aonghusa. Queriocity: A linked data platform for urban information management. In *International Semantic Web Conference*, pp. 148–163. Springer, 2012.

[41] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pp. 3111–3119, 2013.

[42] D. Milne and I. H. Witten. Learning to link with wikipedia. In *Proc. of the 17th ACM conference on Information and knowledge management (CIKM)*, pp. 509–518, 2008.

[43] 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.

[44] Riichiro Mizoguchi. Yamato: yet another more advanced top-level ontology. *Proceedings of the Sixth Australasian Ontology Workshop*, pp. 1–16, 2010.

[45] Abha Moitra, Ravi Palla, Luis Tari, and Mukkai Krishnamoorthy. Semantic inference for pharmacokinetic drug-drug interactions. In *Semantic Computing (ICSC), 2014 IEEE International Conference on*, pp. 92–95. IEEE, 2014.

[46] T. M. Nguyen, T. Kawamura, Y. Tahara, and A. Ohsuga. Self-supervised capturing of users' activities from weblogs. *International Journal of Intelligent Information and Database Systems*, Vol. 6, No. 1, pp. 61–76, 2012.

[47] Ian Niles and Adam Pease. Towards a standard upper ontology. In *Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001*, pp. 2–9. ACM, 2001.

[48] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 701–710. ACM, 2014.

[49] S. F. Pileggi and J. Hunter. An ontological approach to dynamic fine-grained urban indicators. *Procedia Computer Science*, Vol. 108, pp. 2059–2068, 2017.

[50] Thomas Rebele, Fabian Suchanek, Johannes Hoffart, Joanna Biega, Erdal Kuzey, and Gerhard Weikum. Yago: A multilingual knowledge base from wikipedia, wordnet, and geonames. In *International Semantic Web Conference*, pp. 177–185. Springer, 2016.

[51] Petar Ristoski and Heiko Paulheim. Rdf2vec: Rdf graph embeddings for data mining. In *International Semantic Web Conference*, pp. 498–514. Springer, 2016.

[52] Havard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. CRC press, 2005.

[53] H. Santos, V. Dantas, V. Furtado, P. Pinheiro, and D. L. McGuinness. From data to city indicators: A knowledge graph for supporting automatic generation of dashboards. In *Proc. of the 14th Extended Semantic Web Conference (ESWC)*, pp. 94–108, 2017.

[54] Nino Shervashidze, Pascal Schweitzer, Erik Jan van Leeuwen, Kurt Mehlhorn, and Karsten M Borgwardt. Weisfeiler-lehman graph kernels. *Journal of Machine Learning Research*, Vol. 12, No. Sep, pp. 2539–2561, 2011.

[55] Shun Shiramatsu, Teemu Tossavainen, Tadachika Ozono, and Toramatsu Shintani. Towards continuous collaboration on civic tech projects: use cases of a goal sharing system based on linked open data. In *International Conference on Electronic Participation*, pp. 81–92. Springer, 2015.

[56] Evren Sirin, Bijan Parsia, Bernardo Cuenca Grau, Aditya Kalyanpur, and Yarden Katz. Pellet: A practical owl-dl reasoner. *Web Semantics: science, services and agents on the World Wide Web*, Vol. 5, No. 2, pp. 51–53, 2007.

[57] J. F. Sowa. Conceptual graphs for a data base interface. *IBM Journal of Research and Development*, Vol. 20, No. 4, pp. 336–357, July 1976.

[58] John F Sowa. Semantic networks. *Encyclopedia of Cognitive Science*, 2006.

[59] 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.

[60] Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web*, pp. 697–706. ACM, 2007.

[61] Pedro Szekely, Craig A. Knoblock, Jason Slepicka, Andrew Philpot, Amandeep Singh, Chengye Yin, Dipsy Kapoor, Prem Natarajan, Daniel Marcu, Kevin Knight, David Stallard, Subessware S. Karunamoorthy, Rajagopal Bojanapalli, Steven Minton, Brian Amanatullah, Todd Hughes, Mike Tamayo, David Flynt, Rachel Artiss, Shih-Fu Chang, Tao Chen, Gerald Hiebel, and Lidia Ferreira. Building and using a knowledge graph to combat human trafficking. In *Proceedings of the 14th International Semantic Web Conference (ISWC 2015)*, 2015.

[62] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*, pp. 1067–1077. International World Wide Web Conferences Steering Committee, 2015.

[63] Tijmen Tieleman and Geoffrey Hinton. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, Vol. 4, No. 2, pp. 26–31, 2012.

[64] Luis von Ahn. Games with a purpose. *IEEE Computer*, Vol. 39, No. 6, pp. 92–94, 2006.

[65] Denny Vrandečić and Markus Krötzsch. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, Vol. 57, No. 10, pp. 78–85, 2014.

[66] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In *AAAI*, Vol. 14, pp. 1112–1119, 2014.

[67] W. Winkler. The state record linkage and current research problems. *Technical report, Statistics of Income Division, Internal Revenue Service Publication*, 1999.

[68] Ruobing Xie, Zhiyuan Liu, Jia Jia, Huanbo Luan, and Maosong Sun. Representation learning of knowledge graphs with entity descriptions. In *AAAI*, pp. 2659–2665, 2016.

[69] Ruobing Xie, Zhiyuan Liu, Huanbo Luan, and Maosong Sun. Image-embodied knowledge representation learning. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pp. 3140–3146, 2017.

[70] Matthew D Zeiler. Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.

[71] Junbo Zhang, Yu Zheng, and Dekang Qi. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *AAAI*, pp. 1655–1661, 2017.

[72] 芦川将之, 川村隆浩, 大須賀昭彦. クラウドソーシングワーカーの段階的育成方法の提案. 人工知能学会論文誌, Vol. 32, No. 3, pp. B–G81_1, 2017.

[73] 溝口理一郎, 人工知能学会. オントロジー工学. オーム社, 2005.

[74] 堀宗朗, 宮嶋宙, 犬飼洋平, 小国健二. 地震時避難行動予測のためのエージェントシミュレーション. 土木学会論文集 A, Vol. 64, No. 4, pp. 1017–1036, 2008.

[75] 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.

[76] 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.

[77] 來村徳信, 溝口理一郎. 故障オントロジー: 概念抽出とその組織化. 人工知能学会誌, Vol. 14, No. 5, pp. 828–837, 1999.

研究業績

学術雑誌

1. **Shusaku Egami**, Takahiro Kawamura, Kouji Kozaki, Akihiko Ohsuga: Detecting Vicious Cycles in Urban Problem Knowledge Graph using Inference Rules, International Journal on Semantic Web and Information Systems, 採録決定済み, 2018

2. **Shusaku Egami**, Takahiro Kawamura, Kouji Kozaki, Akihiko Ohsuga: Construction of Urban Problem LOD using Crowdsourcing, International Journal of Smart Computing and Artificial Intelligence, Vol.2, No.2, 採録決定済み, 2018

3. Takahiro Kawamura, Katsutaro Watanabe, Naoya Matsumoto, **Shusaku Egami**, Mari Jibu: Funding map using paragraph embedding based on semantic diversity. Scientometrics, Springer, Vol.116, No.2, pp.941-958, 2018.5

4. **Shusaku Egami**, Takahiro Kawamura, Akihiko Ohsuga: Temporal and Spatial Expansion of Urban LOD for Solving Illegally Parked Bicycles in Tokyo. IEICE Transactions on Information and Systems, Vol.E101-D, No.1, pp.116-129, 2018.1

5. **Shusaku Egami**, Takahiro Kawamura, Akihiko Ohsuga: Designing and Publishing Illegally Parked Bicycle LOD. International Journal of Smart Computing and Artificial Intelligence, Vol.1, No.2, pp.77-93, 2017.9

6. **江上周作**，川村隆浩，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，人工知能学会論文誌，30 周年記念論文特集，Vol.31, No.6, pp.AI30-K_1-12, 2016.11

7. **Shusaku Egami**, Takahiro Kawamura, Yuichi Sei, Yasuyuki Tahara, Akihiko Ohsuga: A Solution to Visualize Open Urban Data for Illegally Parked Bicycles. Transactions on Large-Scale Data and Knowledge-Centered Systems XXVII, Springer, pp.129-141, 2016.7

8. **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, 2015.6

国際会議

9. **Shusaku Egami**, Takahiro Kawamura, Akihiko Ohsuga: Predicting Urban Problems: A Comparison of Graph-based and Image-based Methods. Workshop and Poster Proceedings of the 8th Joint International Semantic Technology Conference (JIST2018), pp.114-117, 2018.11

10. Takahiro Kawamura, Katsutaro Watanabe, **Shusaku Egami**, Naoya Matsumoto, Mari Jibu: Content-based Map of Science using Cross-lingual Document Embedding - A Comparison of US-Japan Funded Projects. Proceedings of the 23rd International Conference on Science and Technology Indicators (STI2018), pp.385-394, 2018.10

11. **Shusaku Egami**, Takahiro Kawamura, Kouji Kozaki, Akihiko Ohsuga: Urban Problem LOD for Understanding the Problem Structure and Detecting Vicious Cycles. Proceedings of the 12th IEEE International Conference on Semantic Computing (ICSC2018), pp.186-193, 2018.1

12. Takahiro Kawamura, Katsutaro Watanabe, Naoya Matsumoto, **Shusaku Egami**, Mari Jibu: Science Graph for characterizing the recent scientific landscape using Paragraph Vectors. Proceedings of the 9th International Conference on Knowledge Capture (K-CAP2017), pp.2:1-2:8, 2017.12

13. **Shusaku Egami**, Takahiro Kawamura, Kouji Kozaki, Akihiko Ohsuga: Linked Urban Open Data Including Social Problems' Causality and Their Costs. Proceedings

of the 7th Joint International Semantic Technology Conference (JIST2017), pp.334-349, 2017.11

14. Ryohei Kozu, Takahiro Kawamura, **Shusaku Egami**, Yuichi Sei, Yasuyuki Tahara, Akihiko Ohsuga: User Participatory construction of Open Hazard Data for Preventing Bicycle Accidents. Proceedings of the 7th Joint International Semantic Technology Conference (JIST2017), pp.289-303, 2017.11

15. Takahiro Kawamura, Katsutaro Watanabe, Naoya Matsumoto, **Shusaku Egami**, Mari Jibu: Science Graph for characterizing the recent scientific landscape. Proceedings of the 16th International Semantic Web Conference (ISWC2017) Poster & Demo Track, 2017.10 (ポスター)

16. **Shusaku Egami**, Takahiro Kawamura, Kouji Kozaki, Akihiko Ohsuga: Construction of Linked Urban Problem Data with Causal Relations using Crowdsourcing. Proceedings of the 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI2017), pp.814-819, 2017.7

17. **Shusaku Egami**, Takahiro Kawamura, Akihiko Ohsuga: Estimation of Spatio-temporal Missing Data for Expanding Urban LOD. Proceedings of the 6th Joint International Semantic Technology Conference (JIST2016), pp.152-167, 2016.11

18. Yuri Uehara, Takahiro Kawamura, **Shusaku Egami**, Yuichi Sei, Yasuyuki Tahara, Akihiko Ohsuga: Linked Data Collection and Analysis Platform for Music Information Retrieval. Proceedings of the 6th Joint International Semantic Technology Conference (JIST2016), pp.127-135, 2016.11

19. **Shusaku Egami**, Takahiro Kawamura, Akihiko Ohsuga: Estimation of Spatial Missing Data for Expnading Urban LOD. Workshop and Poster Proceedings of the 6th Joint International Semantic Technology Conference (JIST2016), pp.82-85, 2016.11 (ポスター)

20. Yuri Uehara, Takahiro Kawamura, **Shusaku Egami**, Yuichi Sei, Yasuyuki Tahara, Akihiko Ohsuga: Linked Data Collection and Analysis Platform of Audio Features. Workshop and Poster Proceedings of the 6th Joint International Semantic Technology Conference (JIST2016), pp.78-81, 2016.11 (ポスター)

21. **Shusaku Egami**, Takahiro Kawamura, Akihiko Ohsuga: Building Urban LOD for Solving Illegally Parked Bicycles in Tokyo. Proceedings of the 15th International Semantic Web Conference (ISWC2016), pp.291-307, 2016.10

22. **Shusaku Egami**, Takahiro Kawamura, Yuichi Sei, Yasuyuki Tahara, Akihiko Ohsuga: Schema Design of Illegally Parked Bicycles LOD. Proceedings of the 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI2016), pp.692-697, 2016.7

23. **Shusaku Egami**, Takahiro Kawamura, Yuichi Sei, Yasuyuki Tahara, 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), 2016.1

24. **Shusaku Egami**, Takahiro Kawamura, Akihiro Fujii, Akihiko Ohsuga: Building of Industrial Parts LOD for EDI - A Case Study -. Proceedings of the 4th Joint International Semantic Technology Conference (JIST2014), pp.146-161, 2014.11

25. Akihiro Fujii, **Shusaku Egami**, Hiroyasu Shimizu: EDI support with LOD. Proceedings of the JIST 2013 Joint International Workshop: 2013 Linked Data in Practice Workshop (LDPW2013) and the 1st Workshop on Practical Application of Ontology for Semantic Data Engineering (PAOS2013), pp.27-32, 2013.11

査読付国内シンポジウム

26. 川村隆浩，渡邊勝太郎，松本尚也，**江上周作**，治部眞理: Mapping Science - 飛躍が期待される科学技術領域の抽出 -，第 14 回情報プロフェッショナルシンポジウム

(INFOPRO2017), pp.119-124 2017.11

27. 上原有里，川村隆浩，江上周作，清雄一，田原康之，大須賀昭彦: Linked Data による楽曲特徴量の収集・分析基盤の構築，合同エージェントワークショップ&シンポジウム 2016 (JAWS2016), pp.262-269, 2016.9

28. 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, 2015.10

29. 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, 2014.10

国内大会・研究会

30. 川村隆浩，江上周作，長野伸一，大向一輝，森田武史，山本泰智，古崎晃司: 第 1 回ナレッジグラフ推論チャレンジ 2018〜解釈可能な人工知能を目指して〜，2018 年度人工知能学会全国大会（第 32 回）, 1F1-01, 2018.6

31. 江上周作，川村隆浩，古崎晃司，大須賀昭彦: 都市問題 LOD と推論規則を用いた悪循環の発見，第 44 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-044-08, 2018.3

32. 川村隆浩，渡邊勝太郎，松本尚也，江上周作，治部眞理: Mapping Science - 文書ベクトルを用いた科学技術マップの作成と萌芽領域の抽出 -，研究・イノベーション学会第 32 回年次学術大会予稿集, 2017.10

33. 江上周作，川村隆浩，古崎晃司，大須賀昭彦: クラウドソーシングを用いた社会課題因果関係 LOD の構築，日本ソフトウェア科学会第 34 回大会, MACC3-1-L, 2017.9

34. **江上周作**，川村隆浩，古崎晃司，大須賀昭彦: 社会課題における因果関係を表す Linked Data の半自動的な構築手法の提案，2017 年度 人工知能学会全国大会（第 31 回），1N3-OS-39b-4 , 2017.5

35. **江上周作**，川村隆浩，古崎晃司，大須賀昭彦: 放置自転車問題における因果関係を含んだ LOD の半自動的な構築手法の提案，第 41 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-041-10, 2017.2

36. **江上周作**，川村隆浩，大須賀昭彦: 放置自転車 LOD の拡充に向けた空間的欠損値推定手法の提案，第 39 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-039-06, 2016.9

37. **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, 2016.2

38. **江上周作**，清水宏泰，谷口祥平，藤井章博: ねじ LOD を基にしたマッシュアップアプリケーション，電子情報通信学会人工知能と知識処理研究会, 信学技法 AI2013-17, Vol.113, No.178, pp.13-18, 2013.8

書籍（分担執筆）

39. Takahiro Kawamura, Katsutaro Watanabe, Naoya Matsumoto, **Shusaku Egami**：Chapter 11 - Mapping Science based on research content similarity -, Scientometrics, InTechOpen, ISBN 978-953-51-6239-1, 2018.7

解説論文

40. 川村隆浩，**江上周作**，渡邊勝太郎：研究内容の類似性に基づく科学技術マップの開発，Japio YEAR BOOK 2018 寄稿集, 日本特許情報機構, pp.216-227, 2018.11

41. **江上周作**，渡邊勝太郎：オープンデータのコンテスト型普及活動: LOD チャレンジと JST の連携イベントを例に，情報管理，Vol. 60, No.4, pp.261-270, 2017.7

関連論文の印刷公表の方法及び時期

学術雑誌

1. 全著者名：**<u>Shusaku Egami</u>**, Takahiro Kawamura, Kouji Kozaki, Akihiko Ohsuga
論文題目： Detecting Vicious Cycles in Urban Problem Knowledge Graph using Inference Rules
印刷公表の方法及び時期：International Journal on Semantic Web and Information Systems, 採録決定済み, 2018 年
（第 5 章）

2. 全著者名：**<u>Shusaku Egami</u>**, Takahiro Kawamura, Kouji Kozaki, Akihiko Ohsuga
論文題目：Constructing of Urban Problem LOD using Crowdsourcing
印刷公表の方法及び時期：International Journal of Smart Computing and Artificial Intelligence, Vol.2, No.2, 採録決定済み, 2018 年
（第 5 章）

3. 全著者名：**<u>Shusaku Egami</u>**, Takahiro Kawamura, Akihiko Ohsuga
論文題目：Temporal and Spatial Expansion of Urban LOD for Solving Illegally Parked Bicycles in Tokyo
印刷公表の方法及び時期：IEICE Transactions on Information and Systems, Vol.E101-D, No.1, pp.116-129, 2018 年 1 月
（第 3,4 章）

4. 全著者名：**<u>Shusaku Egami</u>**, Takahiro Kawamura, Akihiko Ohsuga
論文題目：Designing and Publishing Illegally Parked Bicycle LOD
印刷公表の方法及び時期：International Journal of Smart Computing and Artificial

Intelligence, Vol.1, No.2, pp.77-93, 2017 年 9 月
（第 3,4 章）

5. 全著者名：**江上周作**, 川村隆浩, 大須賀昭彦
論文題目：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
印刷公表の方法及び時期：人工知能学会論文誌, 30 周年記念論文特集, Vol.31, No.6, pp.AI30-K_1-12, 2016 年 11 月
（第 3,4 章）

6. 全著者名：**Shusaku Egami**, Takahiro Kawamura, Yuichi Sei, Yasuyuki Tahara, Akihiko Ohsuga
論文題目：A Solution to Visualize Open Urban Data for Illegally Parked Bicycles
印刷公表の方法及び時期：Transactions on Large-Scale Data and Knowledge-Centered Systems XXVII, pp.129-141, 2016 年 7 月
（第 3,4 章）

国際会議

7. 全著者名：**Shusaku Egami**, Takahiro Kawamura, Akihiko Ohsuga
論文題目：Predicting Urban Problems: A Comparison of Graph-based and Image-based Methods
印刷公表の方法及び時期：Workshop and Poster Proceedings of the 8th Joint International Semantic Technology Conference (JIST2018), pp.114-117, CEUR, (ポスター発表), 2018 年 11 月
（第 4 章）

8. 全著者名：**Shusaku Egami**, Takahiro Kawamura, Kouji Kozaki, Akihiko Ohsuga
論文題目：Urban Problem LOD for Understanding the Problem Structure and Detecting Vicious Cycles
印刷公表の方法及び時期：Proceedings of the 12th IEEE International Conference on

Semantic Computing (ICSC2018), pp.186-193, IEEE, 2018 年 1 月
（第 5 章）

9. 全著者名：<u>**Shusaku Egami**</u>, Takahiro Kawamura, Kouji Kozaki, Akihiko Ohsuga
論文題目：Linked Urban Open Data Including Social Problems' Causality and Their Costs
印刷公表の方法及び時期：Proceedings of the 7th Joint International Semantic Technology Conference (JIST2017), pp.334-349, Springer, 2017 年 11 月
（第 5 章）

10. 全著者名：<u>**Shusaku Egami**</u>, Takahiro Kawamura, Kouji Kozaki, Akihiko Ohsuga
論文題目：Construction of Linked Urban Problem Data with Causal RElations using Crowdsourcing
印刷公表の方法及び時期：Proceedings of the 6th IIAI International Congress on Advanced Applied Informatics, (IIAI-AAI2017), pp.814-819, IEEE, 2017 年 7 月
（第 5 章）

11. 全著者名：<u>**Shusaku Egami**</u>, Takahiro Kawamura, Akihiko Ohsuga
論文題目：Estimation of Spatio-temporal Missing Data for Expanding Urban LOD
印刷公表の方法及び時期：Proceedings of the 6th Joint International Semantic Technology Conference (JIST2016), pp.152-167, Springer, 2016 年 11 月
（第 3,4 章）

12. 全著者名：<u>**Shusaku Egami**</u>, Takahiro Kawamura, Akihiko Ohsuga
論文題目：Building Urban LOD for Solving Illegally Parked Bicycles in Tokyo
印刷公表の方法及び時期：Proceedings of the 15th International Semantic Web Conference (ISWC2016), pp.291-307, Springer, 2016 年 10 月
（第 3,4 章）

13. 全著者名：<u>**Shusaku Egami**</u>, Takahiro Kawamura, Yuichi Sei, Yasuyuki Tahara, Akihiko Ohsuga

論文題目：Schema Design of Illegally Parked Bicycles LOD

印刷公表の方法及び時期：Proceedings of the 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI2016), pp.692-697, IEEE, 2016 年 7 月（第 3,4 章）

本論文との関連の詳細

章	関連論文番号	関連する内容
3 章	3, 4, 5, 6, 11, 12, 13	都市問題ごとのナレッジグラフの構築
4 章	3, 4, 5, 6, 7, 11, 12, 13	都市問題ごとのナレッジグラフの時空間的な拡充と分析
5 章	1, 2, 8, 9, 10	都市問題の課題横断的なナレッジグラフの構築と分析

著者略歴

江上 周作（えがみ しゅうさく）

1992 年 1 月 20 日	東京都練馬区に生まれる
2010 年 3 月	錦城高等学校 卒業
2010 年 4 月	法政大学 理工学部 応用情報工学科 入学
2014 年 3 月	法政大学 理工学部 応用情報工学科 卒業
2014 年 4 月	国立大学法人 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 博士前期課程 入学
2016 年 3 月	国立大学法人 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 博士前期課程 修了
2016 年 4 月	国立大学法人 電気通信大学 大学院情報理工学研究科 情報学専攻 博士後期課程 入学
2017 年 6 月	国立研究開発法人 科学技術振興機構 情報企画部 システム担当 派遣社員として就業
2018 年 4 月	独立行政法人 日本学術振興会 特別研究員 (DC2)
2019 年 3 月	国立大学法人 電気通信大学 大学院情報理工学研究科 情報学専攻 博士後期課程 修了予定