

大阪市が抱える社会課題関係情報の Linked Open Data 構築

研究代表者	江上周作	電気通信大学大学院情報理工学研究科 博士後期課程
共同研究者	川村隆浩	科学技術振興財団情報企画部 主任調査員
共同研究者	古崎晃司	大阪大学産業科学研究所 准教授
共同研究者	大須賀昭彦	電気通信大学大学院情報理工学研究科 教授

1 はじめに

現在、国内ではホームレス、放置自転車、ごみ問題、犯罪など様々な社会課題（都市問題）が未解決のまま残っており、2013年G8サミットでの「オープンデータ憲章」の合意以降、行政情報のオープンデータ化と活用によるデータ駆動型の課題解決が注目されている。しかし、社会課題は様々な原因が複雑に絡み合っているため、より課題解決を進めるためには自治体の抱える各種社会課題の分類や因果関係、解決に向けた施策や効果に関するデータの整理が望まれる。

そこで本研究では大阪市を例として、これらのデータのスキーマを整備し、各データの要素をリンクさせた Linked Open Data (LOD) としてデータを蓄積することで、社会課題の分析基盤の構築を目指している。LOD とはデータを主語・述語・目的語の三項構造(トリプル)で記述する Resource Description Framework (RDF) 形式で構造化したオープンデータであり、Web 標準化団体 W3C により推奨されているデータ普及方法である。LOD により誰でも詳細な情報の取得、リンクを辿った横断的な検索やオントロジーに基づいた推論、サービスでの二次利用などが可能になり、データ駆動型の課題解決が促進される。

また、社会課題の本格的な解決には自治体の施策実行が望ましいが、自治体は施策実行に慎重であるため事前に費用対効果の把握を要する。そこで、本研究では大阪市の予算情報を収集し、社会課題の因果関係と紐づけた LOD を構築する。この LOD により、例えば社会課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分析・予測することや、その結果を基に自治体の対策方針や予算削減案を検討することが可能になる。

さらに、本研究で構築した LOD を用いて、社会課題の悪循環および悪循環を引き起こす中核課題を発見することを試みる。複数の社会課題間の因果関係を把握することで、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。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。

2 社会課題関係情報の LOD 構築

2-1 Linked Data スキーマの設計

本研究で構築する LOD 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 1 に設計したスキーマを示す。図 1 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 1 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 Vocabulary¹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

¹ <https://www.w3.org/TR/vocab-data-cube/>

upv: <http://www.ohsuga.lab.uec.ac.jp/urbanproblem/vocabulary#>
 upq: <http://www.ohsuga.lab.uec.ac.jp/urbanproblem/qb#>
 xsd: <http://www.w3.org/2001/XMLSchema>
 qb4o: <http://purl.org/qb4olap/cube#>
 qb: <http://purl.org/linked-data/cube#>

□ Property
 ○ Class

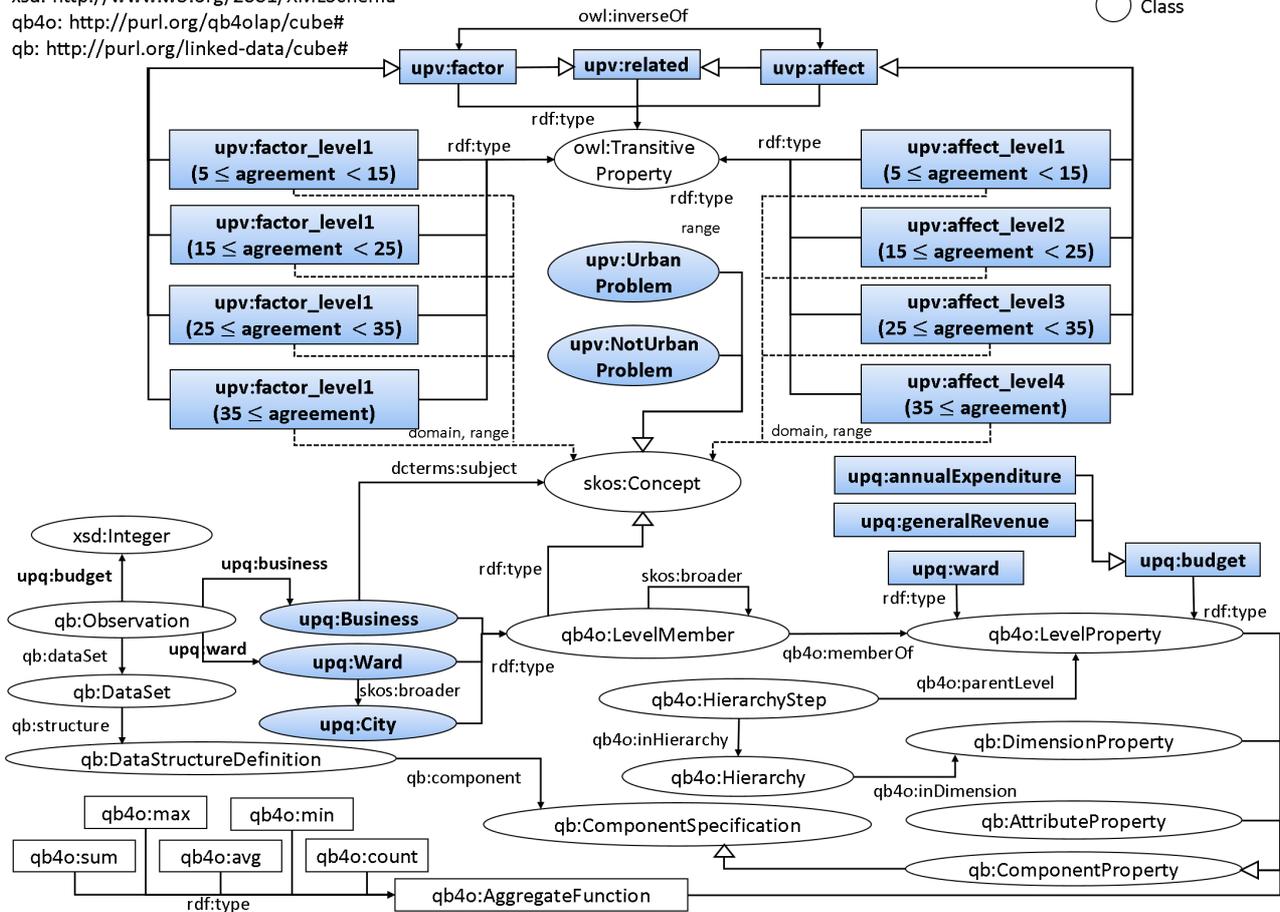


図 1 都市問題の因果関係と予算情報を表す Linked Data スキーマ

拡張である QB4OLAP[1]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 dcterms: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。

2-2 社会課題の因果関係抽出

提案手法は以下の 4 ステップで構成されている。

- (1) 検索エンジンによる 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抽出
- (3) 抽出した因果単語を基にワードクラウドを生成
- (4) クラウドソーシングによる因果単語のフィルタリング

まず、対象とする社会課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 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社会課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

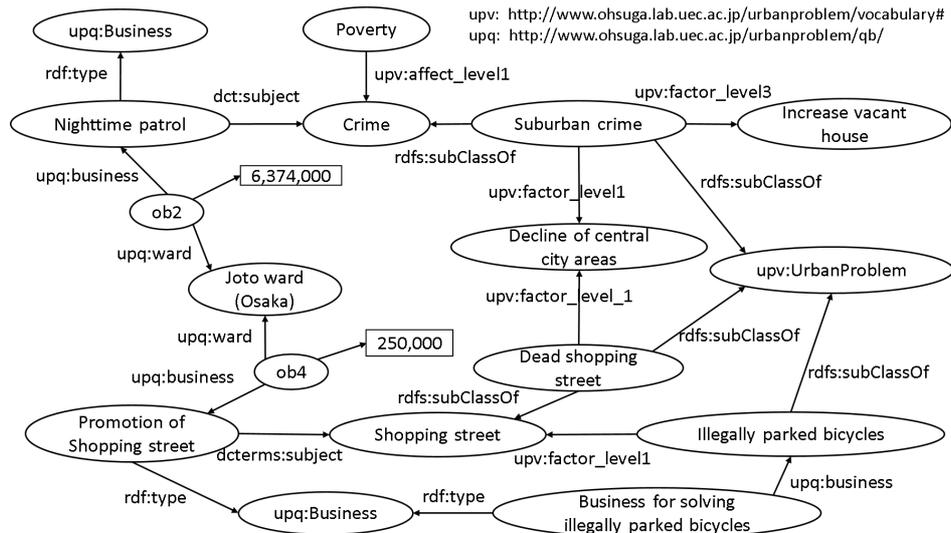


図 3 最終的に構築される LOD の一部

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⁵を使用し、上記 2 つのタスクの遂行報酬を 50 円として、一つの社会課題につき最大 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合は、選択した人数の多い単語に統一する。図 2 内の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、効果などの単語を 20 以上持つ単語のみこのステップを実行する。

2-4 因果単語抽出結果に基づく LOD の構築

2.1 章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 LOD を構築した。クラウドソーシングの結果は TSV データで取得できるため、このデータを基に Apache Jena⁶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 NotUrbanProblem 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 WikiData⁷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 Ontology⁸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 API⁹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。

2-5 大阪市の予算情報に基づく LOD の構築

大阪府は大阪府オープンデータポータルサイトで様々なデータを CC-BY 4.0 ライセンス¹⁰で公開している。本稿では大阪府をモデルとし、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、

⁵ <https://www.lancers.jp/>

⁶ <https://jena.apache.org/>

⁷ https://www.wikidata.org/wiki/Wikidata:Main_Page

⁸ <https://www.w3.org/TR/prov-o/>

⁹ <https://cloud.google.com/translate/>

¹⁰ <https://creativecommons.org/licenses/by/4.0/deed.ja>

表 1 因果単語抽出結果

	# of documents including urban problem words	# of sentences including synonyms of “factor” and “influence”	# of extracted words
Factor	1,438	4,481	3,110
Influence	2,465	9,082	4,661

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 POI¹¹を用いてセルから情報を抽出し、Apache Jena を用いて 2.1 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の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ことで、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 `dc:subject` プロパティでリンク付けを行う。

図 3 は最終的に構築される LOD の一部分である。トリプル数は 49,386 となった。構築した LOD は我々のウェブサイト¹²で、LOD に対して検索が実行可能な SPARQL Protocol and RDF Query Language (SPARQL)¹³ エンドポイントとともに公開している。

2-6 因果関係抽出結果の考察

表 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。2.2 章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 [3] を計算した。ワーカー数は 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。従って、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、本手法はある程度妥当であったといえる。一致度が下がった原因として、社会課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機能しない例があったことにより生じている。また、ワーカーの予備知識が共通していないことも考えられ、社会課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、社会課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、ワーカーの予備知識が共通していたのではないかと考える。

3 社会課題解決施策の検討に向けた LOD の検索

本章では構築した LOD を活用し、社会課題の因果関係から関連する予算情報まで検索する例を紹介する。

¹¹ <https://poi.apache.org/>

¹² <http://www.ohsuga.is.uec.ac.jp/urbanproblem/>

¹³ <https://www.w3.org/TR/rdf-sparql-protocol/>

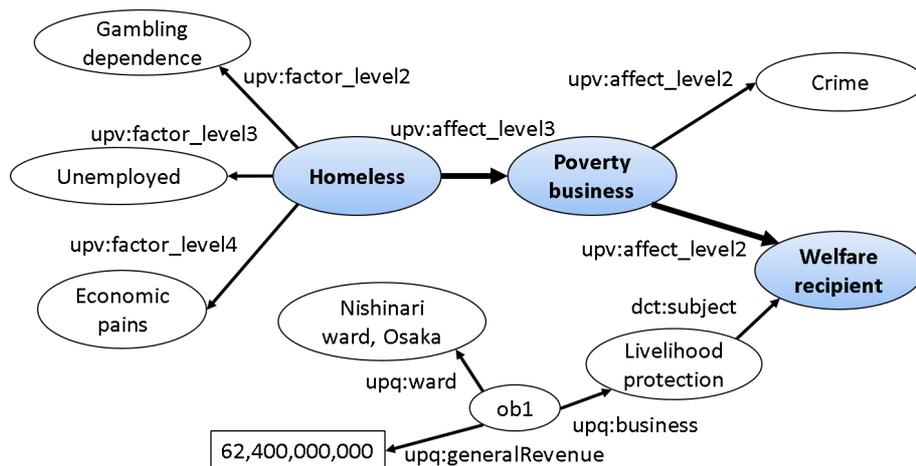


図 4 ホームレスの因果関係

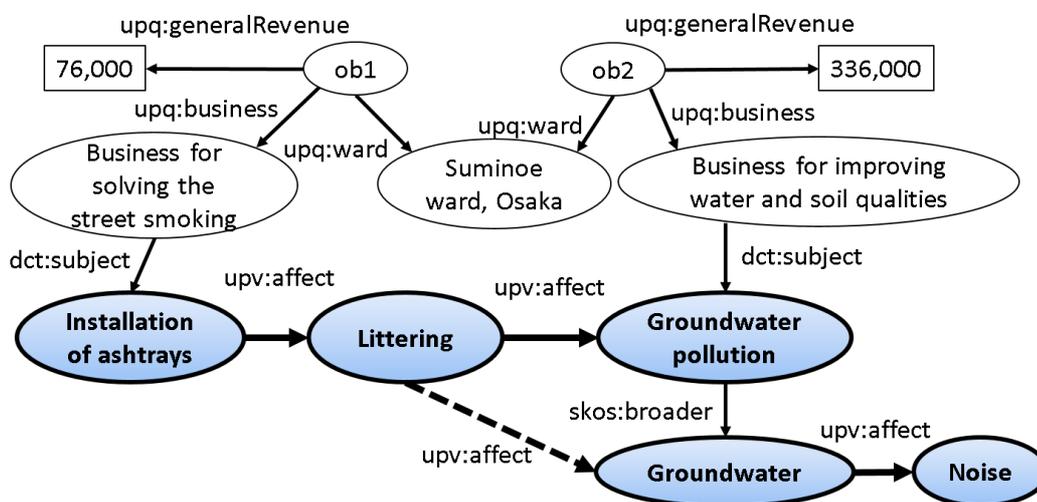


図 5 ポイ捨ての因果関係

3-1 ホームレスの因果関係

図 4 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事がわかる。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。

2015 年の統計データ¹⁴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。

¹⁴ <http://www.city.osaka.lg.jp/fukushi/page/0000197783.html>

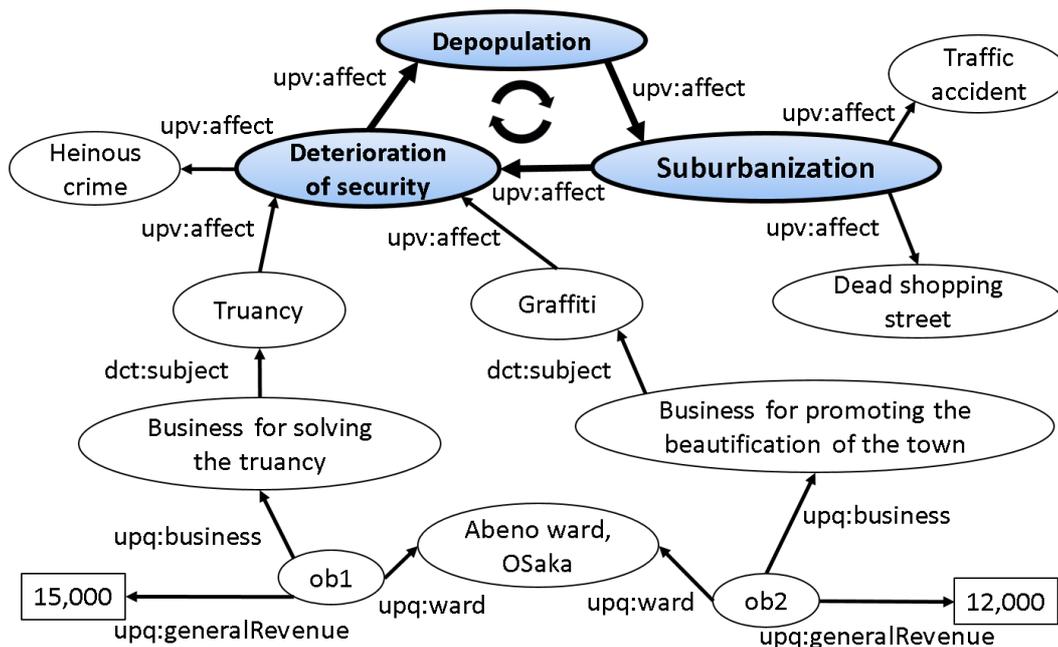


図 6 郊外化の因果関係

3-2 ポイ捨ての因果関係

図 5 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、4 章で後述する、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ルールにより、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。

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壤汚染・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、土壤汚染・水質汚濁改善事業は予算を削減できる可能性がある。

3-3 郊外化の因果関係

図 6 は郊外化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、「郊外化→治安悪化→人口減少→郊外化」の悪循環が確認できる。郊外化（中心市街地の空洞化）により人通りの減少や空き家や増加することで治安が悪化する。治安悪化により犯罪が多発することでその地域からの人口流出、人口減少が加速し、さらなる郊外化（中心市街地の空洞化）に繋がる。また、郊外化に関するこの悪循環が繰り返される過程で、交通事故、シャッター通り、凶悪犯罪の発生などの他の問題に影響を与え、これらの二次的被害が拡大する恐れがあることがわかる。したがって、それ自体が深刻化するだけでなく、他の都市問題を次々と発生させる郊外化の悪循環は、対策が急がれる問題であると言える。

一方で郊外化の悪循環に突入する要因として不就学や落書きがある。不就学に関しては、素行不良の青少年が増加することが治安悪化につながると考えられる。落書きに関しては、落書きを放置しておくことで自

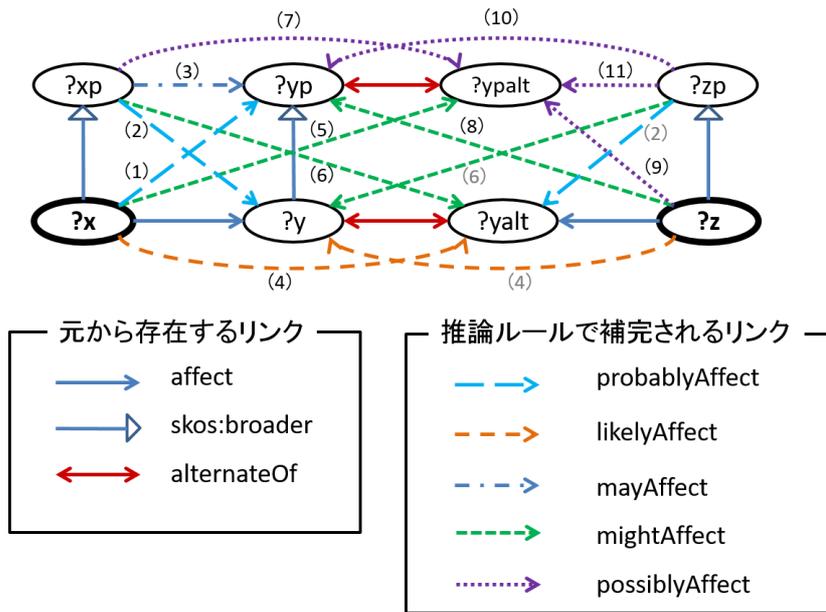


図 7 推論ルールにより補完される欠損因果リンク

治体が機能していないというイメージを与え、治安の悪化につながると考えられる。これは割れ窓理論[4]に相当する現象である。しかしながら、大阪市阿倍野区において就学対策費は 15,000 円、まちの美化推進事業の予算は 12,000 円と極めて低く、大阪市全体で見ても就学対策費は平均 15,083 円、まちの美化推進事業の予算は平均 14,000 円とあまり重要視されていないことが伺える。これらの問題への対策予算を増加することで、郊外化の悪循環に入り込むリスクを減らすことが出来ると考えられる。

本章で紹介した因果関係と大阪市予算の関係から見る課題解決策検討例は、区・シティマネージャー自由経費予算に基づく一例である。大阪市は他にも様々な事業予算があるため、多角的に課題解決策を検討できるよう、今後より多くの予算情報を LOD に追加し拡充する必要がある。

4 都市問題の悪循環発見手法

3-3 節に見られるように都市問題は悪循環する傾向にある。本章では、都市問題の悪循環構造や中核課題を把握するため、構築した LOD からそれらを発見する手法について説明する。

4-1 推論ルールを用いた因果リンクの欠損補完

構築した LOD には UrbanProblem クラス間や CausalEntity (NotUrbanProblem) 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 情報から半自動的に構築しているため、実際に起こりうる全ての因果関係が含まれているわけではなく欠損が生じている。そこで、独自に定義した推論ルールにより欠損リンクを補完した。図 4 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 7 中の数字は次の Semantic Web Rule Language (SWRL)¹⁵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) $(?x \text{ upv:affect } ?y), (?y \text{ skos:broader } ?yp) \rightarrow (?x \text{ upv:probablyAffect } ?yp) .$
- (2) $(?x \text{ upv:affect } ?y), (?x \text{ skos:broader } ?xp) \rightarrow (?xp \text{ upv:probablyAffect } ?y) .$
- (3) $(?x \text{ upv:affect } ?y), (?x \text{ skos:broader } ?xp), (?y \text{ skos:broader } ?yp) \rightarrow (?xp \text{ upv:mayAffect } ?yp) .$
- (4) $(?x \text{ upv:affect } ?y), (?y \text{ prov:alternateOf } ?yalt) \rightarrow (?x \text{ upv:likelyAffect } ?yalt) .$
- (5) $(?x \text{ upv:affect } ?y), (?y \text{ skos:broader } ?yp), (?yp \text{ prov:alternateOf } ?ypalt) \rightarrow (?x \text{ upv:mightAffect } ?ypalt) .$
- (6) $(?x \text{ upv:affect } ?y), (?x \text{ skos:broader } ?xp), (?y \text{ prov:alternateOf } ?yalt) \rightarrow (?xp \text{ upv:possiblyAffect } ?yalt) .$

¹⁵ <https://www.w3.org/Submission/SWRL/>

```

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 skos:broader ?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)
}

```

図 8 都市問題の悪循環発見のための SPARQL クエリ

表 2 発見された悪循環の数

Vicious cycles	Normal	Inference
3 nodes	33	45
4 nodes	168	308
5 nodes	236	460
6 nodes	514	1,091
Total	951	1,904

upv:mightAffect ?yalt) .

(7) (?x upv:affect ?y), (?x skos:broader ?xp), (?y skos:broader ?yp), (? yprov: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) .

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 Stardog¹⁶に格納している. Stardog は OWL 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 で記述して格納している. 結果として, 1,058 の probablyAffect プロパティ, 122 の likelyAffect プロパティ, 191 の mayAffect プロパティ, 333 の mightAffect プロパティ, 179 の possiblyAffect プロパティが推論された.

¹⁶ <https://www.stardog.com/>

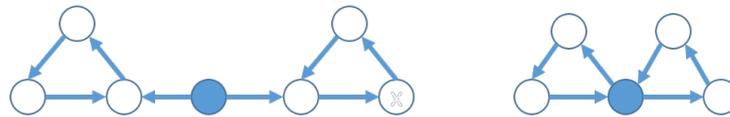
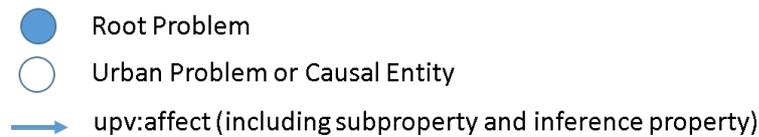


図9 根源的課題のグラフパターン

4-2 都市問題の悪循環の発見

次に、SPARQL を用いて LOD から都市問題の因果関係の悪循環を発見する。本稿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。図8は都市問題の悪循環発見のためのSPARQLクエリであり、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 -affect-> Truancy -affect-> Diseaseと、Truancy -affect-> Disease -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2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。

4-3 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。本稿では図9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9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、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る。悪循環のサイズは3ノードから6ノードまで変化する。これらの根源的課題をSPARQLクエリにより発見する。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。

5 評価と考察

5-1 発見した悪循環の評価と考察

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、本稿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ほー無レスの一員となっていることは確かであるため、多重債務-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ごみ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-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-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-affect->貧困ビジネス-affect->日雇い労働-affect->ホームレス-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 発見した根源的課題の考察

4.3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-affect->ホームレス-affect->貧困ビジネス-affect->日雇労働の正解悪循環と治安悪化-affect->落書き-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思われる。

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、都市問題 LOD は課題解決に向けた専門家同士の議論活性化ツールとして有用であるとのコメントを頂いた。一方で、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[5]等の適用を検討している。

6 むすび

本研究では大阪市が抱える社会課題関係情報を記述する LOD を構築するため、社会課題の因果関係とその予算を記述する Linked Data スキーマを設計し、自然言語処理とクラウドソーシングを用いて因果関係の抽出を行った。また、構築した LOD 内の欠損リンクを推論ルールにより補完し、SPARQL を使用して都市問題の悪循環および悪循環につながる根源的課題を発見する手法を提案した。さらに、発見した悪循環の起こりうる可能性を専門家の意見を基に評価した。本研究で構築した LOD の活用が、大阪市の抱える都市問題の構造把握や課題発見、解決策検討に向けた議論促進等に有用なことを示した。今後は冗長的な表現の多い文章の平易化や、機械学習等を用いた因果関係の抽出により、悪循環発見の精度向上を試みる。また、LOD 内の因果関係に数値情報を付加することについても検討している。

本研究では社会課題の因果関係は行政による分析、社会学分野の研究成果、ソーシャルデータ、事実、人間の知識や連想などの情報に基づいて構築されている。従って、これらのソースデータを辿ることで全ての因果関係を説明することが可能である。今後、因果関係の理解が困難な場合に中間ノードを増やすことを検討している。

【参考文献】

- [1] Etcheverry, L., Vaisman, A., and Zimányi, E.: Modeling and Querying Data Warehouses on the Semantic Web using QB4OLAP, Proc. of the 16th International Conference on Data Warehousing and Knowledge Discovery (DaWaK), 2014, pp.45-56.
- [2] Winkler, W.: The state record linkage and current research problems, Technical report, Statistics of Income Division, Internal Revenue Service Publication, 1999.
- [3] J. L. Fleiss, J. Cohen, "The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability," Educational and Psychological Measurement, Vol.33, No.3, 1973, pp.613-619
- [4] Wilson, J. Q. and K., G. L.: Broken windows, Critical issues in policing, 1982, pp. 395-407.
- [5] F. Bobillo and U. Straccia "Fuzzy ontology representation using OWL 2," International Journal of Approximate Reasoning, Vol. 52, No. 7, 2011, pp.1073-1094

〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
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	2018 年 3 月
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 年 1 月
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 年 1 月
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 年 11 月
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	2017 年 9 月
Designing and Publishing Illegally Parked Bicycle LOD	International Journal of Smart Computing and Artificial Intelligence, Vol.1, No.2, pp. 77-93	2017 年 9 月
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 年 7 月
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)	2017 年 5 月