

## 科学研究費助成事業 研究成果報告書

令和元年6月5日現在

機関番号：12612

研究種目：基盤研究(C)（一般）

研究期間：2016～2018

課題番号：16K00419

研究課題名（和文）学術情報の意味的・計量的分析に向けた論文抄録のリンクトデータ化手法に関する研究

研究課題名（英文）Studies on triplification methods of paper abstracts for semantical and quantitative scientific information analysis

研究代表者

川村 隆浩（Kawaumra, Takahiro）

電気通信大学・大学院情報理工学研究科・客員准教授

研究者番号：10426653

交付決定額（研究期間全体）：（直接経費） 3,400,000円

研究成果の概要（和文）：本研究では、学術情報を意味的、計量的に分析するため、主に学術論文の抄録文を自然言語処理と機械学習技術によってLinked Data形式に変換する手法の研究・開発を実施した。

Linked Dataとは、情報の内容を主語-述語-目的語などの組で表現するフォーマットであり、国際的に標準化されたものである。

専門的な内容を日本語、英語の言語の壁も超えて統一的なフォーマットで表すことで、科学技術情報を内容的にも数値的にも客観的に概観、分析することを可能にした。

研究成果の学術的意義や社会的意義

近年、日本の科学技術力の低下が指摘される中、論文や特許、ファンディングプロジェクト情報といった学術情報の1次ソースを客観的に分析できる基盤（内容に基づく科学技術マップ）を構築した。

本マップ上でさまざまな検索や分類を行うことで、日本が強い分野や弱い分野、米国や欧州において資金を注いでいる分野、およびそれらに関わる研究者の一覧や関係性など、日本と海外の比較を行うことができる。

それによって、現在の日本の立ち位置を測り、今後の日本の科学技術の方向性を検討するのに役立つだろう。

研究成果の概要（英文）：This study focused on the triplification methods, which translates scientific paper abstracts, etc. to Linked Data using natural language processing and machine learning techniques for semantically and quantitatively analyzing scientific inform.

Linked Data is an internationally-standardized format, in which information contents are represented by triple, i.e., subject - verb - object.

The unified format that can represent professional contents written in English and Japanese helps us to objectively browse and analyze the scientific and technological information in contents and in metrics.

研究分野：人工知能，学術情報分析

キーワード：ナレッジグラフ Linked Data 知識処理

## 1. 研究開始当初の背景

学術情報分析の分野では、以前より発行された論文の引用数に基づく学術的な評価、分析が盛んに行われていた。しかし、論文が一定の回数引用されるには数年単位の時間がかかる、国際会議論文は相対的に引用数が少ない傾向がある、などいくつかの問題が指摘されている。一方、論文の内容に基づく分析はTF・IDFなど一部の手法に留まっていた。

## 2. 研究の目的

そこで本研究では、学術情報を意味的、計量的に分析するため、主に学術論文の抄録文を自然言語処理と機械学習技術によってLinked Data形式に変換する手法の研究・開発を実施した。

Linked Dataとは、情報をW3C勧告であるRDFデータモデルで構造化し、情報間の意味的な関係を相互にリンクしたものである。2015時点において、欧米を中心に630億トリプル(RDBにおけるレコードに相当)がWeb上に公開されており、研究データの標準的なフォーマットとして広がりつつある。

## 3. 研究の方法

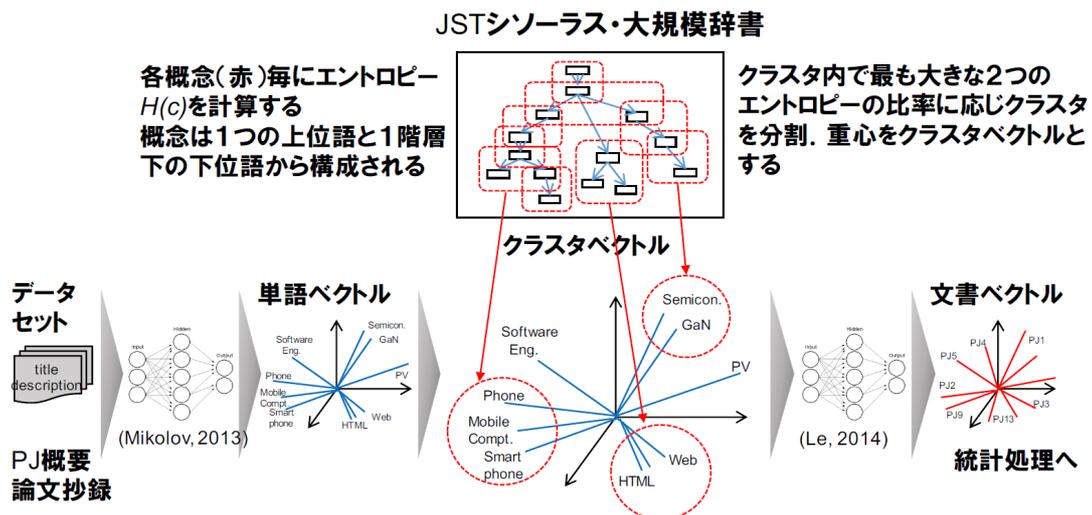
具体的には、1. 主に学術論文の抄録文を対象にLinked Dataを構成する要素(主語、述語、目的語、場所、時間など)を抽出し、2. 抽出語句を既存Linked Dataやオントロジーへマッピングする手法の研究開発を行った。これにより、英語や日本といった自然言語で記述された非構造化データであった抄録情報の意味的、計量的な分析を可能にするものである。

## 4. 研究成果

まず、論文や特許、ファンディングプロジェクト情報といった自然言語で書かれた情報を、科学技術情報に特化してベクトル化する手法を研究開発した。

# 文書ベクトル化手法

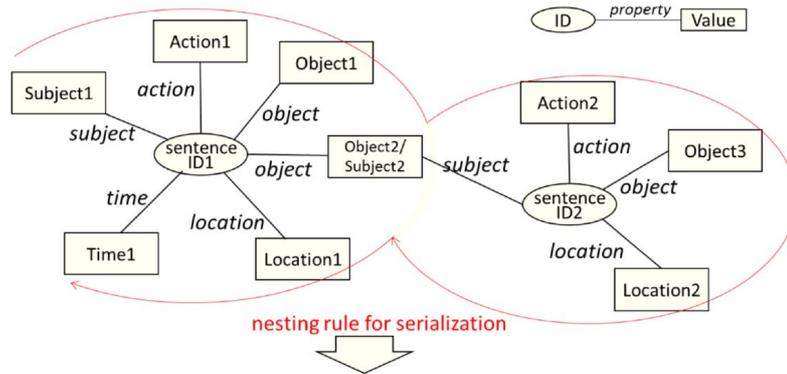
- word2vecを用いて単語ベクトルを生成し、単語ベクトルを情報エンタロピーに基づいてクラスタリングし、文書ベクトルを生成する手法を提案 [Kawamura'18]



また、それらを Linked Data 形式に変換する手法を研究開発した。

# 言語横断的な文書ベクトル化手法

- 国内外のプロジェクトを同時にマップ上で分析するために, Semantic Role Labeling (SRL, 意味的役割付与) に基づく文章のトリプル化手法[川村'16]によりSRグラフを生成し, シリアライズしてから文書ベクトルを生成する手法を開発[Kawamura'18]

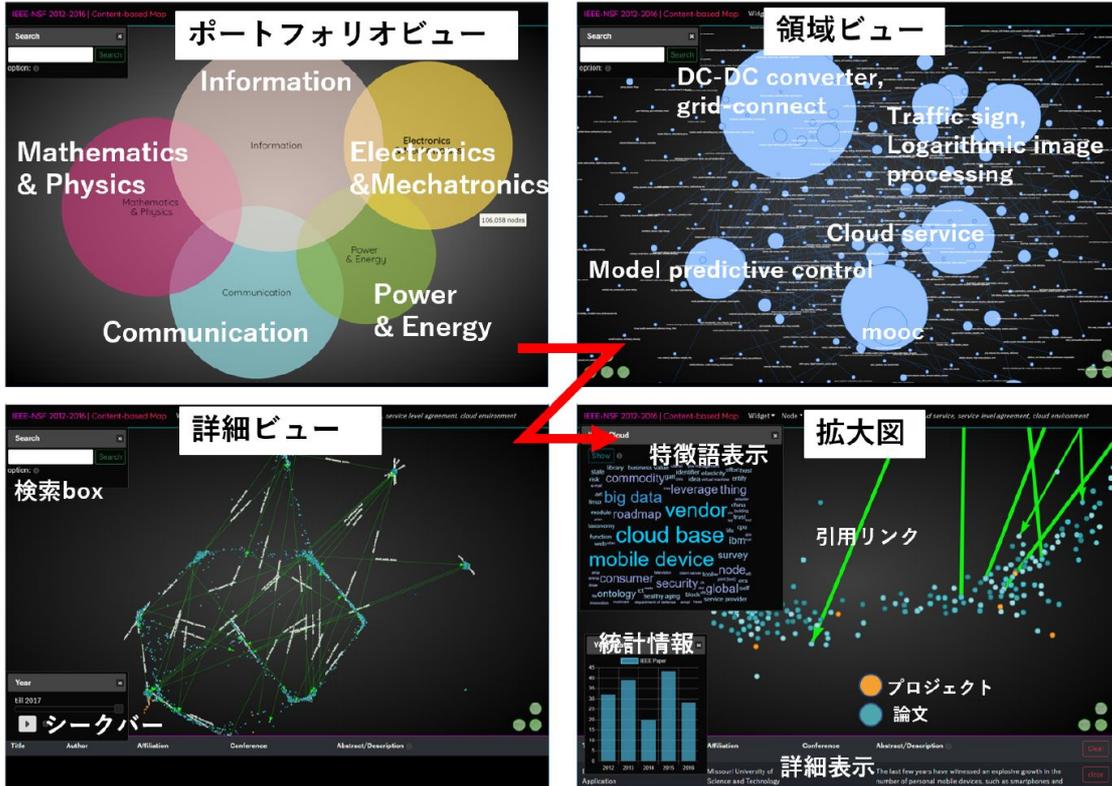


{Subject1 Action1 Object1 Object2/{Subject2 Action2 Object3 Location2} Location1 Time1}

[Kawamura'18] Kawamura, T., Watanabe, K., Egami, S., Matsumoto, N., and Jibu, M. Content-based map of science using crosslingual document embedding—a comparison of us/japan funded projects. In 23rd International Conference on Science and Technology Indicators (STI 2018), pp. 385–394, 2018.

[川村'16] 川村隆浩, 大須賀昭彦. Text2lod ~テキスト情報の lod化に向けた web api の開発~. 人工知能学 会論文誌, Vol. 31, 7 No. 1, pp. 1–8, 2016.

さらに, これらの情報を客観的に分析できる基盤 (内容に基づく科学技術マップ) を構築した.



[https://jipsti.jst.go.jp/foresight/mapping\\_science/](https://jipsti.jst.go.jp/foresight/mapping_science/)

このマップ上でさまざまな関係性を辿った高度検索や, 推論処理, 統計的な分類やクラスタリングを行うことで, 日本と海外の学術情報やファンディングプロジェクトの比較を行うことができ, 今後の日本の科学技術政策の検討に役立つことを期待している.

## 5. 主な発表論文等

### [雑誌論文](計6件)

T. Kawamura, K. Watanabe, N. Matsumoto, S. Egami, M. Jibu: Funding Map using Paragraph Embedding based on Semantic Diversity, *Scientometrics*, Springer, Vol. 116, No. 2, pp. 941-958, 2018. <https://doi.org/10.1007/s11192-018-2783-x> (5-Year Impact Factor:2.346) 査読有

S. Egami, T. Kawamura, K. Kozaki, A. Ohsuga: Construction of Urban Problem LOD using Crowdsourcing, *International Journal of Smart Computing and Artificial Intelligence*, Vol. 2, No. 1, pp. 79-94, 2018. 査読有

S. Egami, T. Kawamura, A. Ohsuga: Temporal and Spatial Expansion of Urban LOD for Solving Illegally Parked Bicycles in Tokyo, *IEICE Transaction on Information & Systems*, Vol. E101-D, No. 1, pp. 116-129, 2018. 査読有

S. Egami, T. Kawamura, A. Ohsuga: Designing and Publishing Illegally Parked Bicycle LOD, *International Journal of Smart Computing and Artificial Intelligence*, Vol. 1, No. 2, pp. 77-93, 2017. 査読有

T. Kawamura, M. Sekine, K. Matsumura: Detecting Hypernym/Hyponym in Science and Technology Thesaurus using Entropy-based Clustering of Word Vectors, *International Journal of Semantic Computing, Special Issue on the best papers of IEEE ICSC 2017*, Vol. 11, No. 4, pp. 433-449, 2017. 査読有

江上 周作, 川村 隆浩, 清 雄一, 田原 康之, 大須賀 昭彦: 社会課題解決に向けた循環型 LOD 構築システムの提案, *人工知能学会論文誌 30 周年記念論文特集*, Vol. 31, No. 6, pp. A130-K\_1-12, 2017. 査読有

### [学会発表](計15件)

江上 周作, 渡邊 勝太郎, 川村 隆浩: Mapping Science - ナレッジグラフに基づく科学技術マップの高度検索と対話的操作の実現, *人工知能学会セマンティック Web とオントロジー研究会(SIG SWO)*, *人工知能学会研究会資料*, SIG-SWO-047-13, 2019. (人工知能学会研究会優秀賞受賞)

川村 隆浩: トーク&トーク「AI とインフォプロの将来」, 第 15 回情報プロフェッショナルシンポジウム (INFOPRO2018), 2018. (招待講演)

T. Kawamura, K. Watanabe, N. Matsumoto, S. Egami, M. Jibu: Content-based Map of Science using Cross-lingual Document Embedding, *Proceedings of 23rd International Conference on Science and Technology Indicators (STI 2018)*, pp.385-394, 2018. 査読有

S. Egami, T. Kawamura, K. Kozaki, A. Ohsuga: Urban Problem LOD for Understanding the Problem Structure and Detecting Vicious Cycles, *Proceedings of 12th IEEE International Conference on Semantic Computing (ICSC 2018)*, pp.186-193, 2018. 査読有

T. Kawamura, K. Watanabe, N. Matsumoto, S. Egami, M. Jibu: Science Graph for characterizing the recent scientific landscape using Paragraph Vectors, *Proceedings of 9th ACM International Conference on Knowledge Capture (K-Cap 2017)*, pp. 9-16, 2017. 査読有

S. Egami, T. Kawamura, K. Kozaki, A. Ohsuga: Linked Urban Open Data Including Social Problems' Causality and Their Costs, *Proceedings of 7th Joint International Semantic Technology Conference (JIST 2017)*, LNCS 10675, Springer, pp. 334-349, 2017. 査読有

R. Kozu, T. Kawamura, S. Egami, Y. Sei, Y. Tahara, A. Ohsuga: Building of User Participatory Open Hazard Data for Preventing of Bicycle Accidents, *Proceedings of 7th Joint International Semantic Technology Conference (JIST 2017)*, LNCS 10675, Springer, pp. 289-303, 2017. 査読有

T. Kawamura, K. Watanabe, N. Matsumoto, M. Jibu: Funding Map for Research Project Relationships using Paragraph Vectors, *Proceedings of 16th International Conference on Scientometrics & Informetrics (ISSI 2017)*, pp. 1121-1131, 2017. 査読有

S. Egami, T. Kawamura, K. Kozaki, A. Ohsuga: Construction of Linked Urban Problem Data with Causal Relations using Crowdsourcing, *Proceedings of 6th IIAI International Congress on Advanced Applied Informatics (AAI 2017)*, pp. 814-819, 2017. 査読有

T. Kawamura, M. Sekine, K. Matsumura: Hyponym/Hypernym Detection in Science and Technology Thesauri from Bibliographic Datasets, *11th IEEE International Conference on Semantic Computing (ICSC 2017)*, Poster session, 2017. 査読有

T. Kawamura, K. Kozaki, T. Kushida, K. Watanabe, K. Matsumura: Expanding Science and Technology Thesauri from Bibliographic Datasets using Word Embedding, *Proceedings of 28th IEEE International Conference on Tools with Artificial Intelligence (ICTAI*

2016), pp. 857-864, 2015. 査読有

S. Egami, T. Kawamura, A. Ohsuga: Estimation of Spatio-temporal Missing Data for Expanding Urban LOD, Proceedings of 6th Joint International Semantic Technology Conference (JIST 2016), LNCS 10055, Springer, pp. 152-167, 2016. 査読有

Y. Uehara, T. Kawamura, Y. Sei, Y. Tahara, A. Ohsuga: Linked Data Collection and Analysis Platform for Music Information Retrieval, Proceedings of 6th Joint International Semantic Technology Conference (JIST 2016), LNCS 10055, Springer, pp. 127-135, 2016. 査読有

S. Egami, T. Kawamura, A. Ohsuga: Building Urban LOD for Solving Illegally Parked Bicycles in Tokyo, Proceedings of 14th International Semantic Web Conference (ISWC 2016), LNCS 9982, Part II, Springer, pp. 291-307, 2016. 査読有

S. Egami, T. Kawamura, A. Ohsuga: Schema Design of Illegally Parked Bicycles LOD, Proceedings of 5th IIAI International Congress on Advanced Applied Informatics (AAI 2016), pp. 692-697, 2016. 査読有

[ 図書 ] (計 3 件)

伊神 正貴, 村上 昭義, 川村 隆浩: サイエンスマップ 2016 - 論文データベース分析 (2011-2016 年)による注目される研究領域の動向調査, 文部科学省科学技術・学術政策研究所, 142 頁, 2018. (共著)

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, 特許×AI 特集, Japio YEAR BOOK 2018, 一般財団法人 日本特許情報機構, pp. 80-91, 2018. (分担執筆)

T. Kawamura, K. Watanabe, N. Matsumoto, S. Egami, M. Jibu: Chapter 11. Mapping Science based on research content similarity, Scientometrics, InTechOpen, pp. 175-194, ISBN:978-953-51-6239-1, 2018. (分担執筆)

[ その他 ]

Mapping Science (サイエンスマップ),

[https://jipsti.jst.go.jp/foresight/mapping\\_science/](https://jipsti.jst.go.jp/foresight/mapping_science/)

6 . 研究組織

(1)研究分担者

研究分担者氏名: 大須賀 昭彦

ローマ字氏名: Ohsuga Akihiko

所属研究機関名: 電気通信大学

部局名: 大学院情報理工学研究科

職名: 教授

研究者番号 (8 桁): 90393842

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。