

社会課題解決に向けたナレッジグラフと欠損推定手法の提案

～学内駐輪環境改善の試み～

非会員 塚越 雄登^{*a)} 非会員 川村 隆浩^{**} 非会員 清 雄一^{*}
非会員 田原 康之^{*} 正員 大須賀昭彦^{*}

Proposal of Knowledge Graph and Completion Method for Solving Social Issues: An Attempt to Improve the Bicycle Parking Environment in the University

Yuto Tsukagoshi^{*a)}, Non-member, Takahiro Kawamura^{**}, Non-member, Yuichi Sei^{*}, Non-member, Yasuyuki Tahara^{*}, Non-member, Akihiko Ohsuga^{*}, Member

(2019年10月29日受付, 2020年5月15日再受付)

Societies in Japan face many social issues today. To address these issues, the government is exploring ways to solve problems using data. The purpose of this study is to provide support using existing data for grasping the current situation, taking up the problem of bicycle parking environment in a university as an example of social issues. Specifically, we tried to design a unique schema and build knowledge graph for integration of various on-campus data. Then we applied and modified the knowledge graph completion methods to improve efficiency and accuracy. For the knowledge graph constructed in this study, the number of bicycles was estimated by the conventional method and the proposed method, and the transition of the sum of absolute value errors was compared. The proposed method exceeded the existing method in efficiency and accuracy. In addition, approximately 650 data aggregated at any date and place were able to be estimated about 54 units correctly, compared to the conventional method. Finally, we improved efficiency and accuracy of knowledge graph completion methods.

キーワード: ナレッジグラフ, Linked open data, 欠損推定, Translation-based model

Keywords: knowledge graph, Linked open data, knowledge graph completion, Translation-based model

1. はじめに

今日、我が国には放置自転車問題、ごみ問題、ホームレス問題、インフラ老朽化問題など様々な社会課題が存在する。これらは多様な要因が重なった結果生じている問題であり、当該事例のみに関わる単一的な視点での現状把握や解決策によって解消することは難しい。しかし現段階ではまだ、複合的な観点を考慮した課題解決の例は少ないのが

現状である。

内閣府^①は、このような社会課題に対してデジタル技術を徹底的に活用した解決策を見出すことが、大きな意義を持つと考えている。また、デジタル時代における「新たな資源」であるデータに関しては、一定の条件下で自由にデータが流通する環境を整備し、国民生活で便益を実感できるデータを利活用したイノベーションを促進するとしており、メーカー・ベンダーの枠を超えたデータ連携を急いでいる。

このように複数データの連携や、それらを公開して利活用することによる社会課題解決が注目されており、総務省も、公共データのオープンデータ化が、新事業・新サービスの創出、行政サービスの高度化、地域の経済活性化・課題解決等に寄与するもとであるとし、大きな可能性に言及している^②。

一方、行政機関や民間事業者等においては、多種類のデータ形式でデータ生成、保存を行っており、データ連携は容易ではない上に、データ欠損なども妨げとなっている^③。これらの理由から、現状日本国内では再利用性に欠けるデー

a) Correspondence to: Yuto Tsukagoshi. E-mail: tsukagoshi.yuto@ohsuga.lab.uec.ac.jp

* 電気通信大学大学院 情報理工学研究科
〒182-8585 東京都調布市調布ヶ丘 1-5-1

Graduate School of Informatics and Engineering,
The University of Electro-Communications
1-5-1, Chofugaoka, Chofu, Tokyo 182-8585, Japan

** 農業・食品産業技術総合研究機構

〒305-8517 茨城県つくば市観音台 3-1-1

National Agriculture and Food Research Organization
3-1-1, Kannondai, Tsukuba, Ibaraki 305-8517, Japan

タが多く、データを活用した社会課題解決は十分に進んでいない。

また、分野によってデータ標準化やデータ連携の進捗状況に差があるため、現時点では分野横断的なデータ連携を見据えつつ、政府や地方公共団体などが保有するデータの集積を進めるとともに、各分野のデータ標準化やデータ連携を進めることも重要であるとされている⁽¹⁾。

ここで、データ形式の統一やオープンデータの有用性向上のために、Linked Open Data(LOD)が推奨されている⁽⁴⁾。LODはオープンデータとLinked Dataの融合であり、オープンであると同時に互いにリンクされている⁽⁵⁾。ここでいうLinked Dataとは、Resource Description Framework(RDF)を用いて構造化される意味的な繋がりを持つデータであり、広義にはナレッジグラフともよばれる。

社会課題に関するデータを、このナレッジグラフ上に統合してweb上に公開することで、誰でも社会課題解決に向けたデータの利活用が可能になる。また、ナレッジグラフ化することにより、従来のように一つの社会課題単体ではなく、それと繋がるあらゆるデータを鑑みた複合的な観点から解決の糸口を探ることが容易になる。

また、Ehlingerら⁽⁶⁾の定義によれば、ナレッジグラフとは、獲得した知識をオントロジーを使って統合し、推論を適用することで新しい知識を導き出すものである。よって、ナレッジグラフには、データの統合・再構築だけでなく、欠損推定の能力活用が期待されている。

近年、このナレッジグラフにおける意味的な繋がりを利用した欠損推定手法の研究が盛んに行われており、H. Yoonら⁽⁷⁾によれば、中でも特にグラフを低次元の連続ベクトル空間に埋め込む手法は最も有望な方法の一つであり、同時に、TransE⁽⁸⁾が最もよく知られた手法である。H. Moussellyら⁽⁹⁾によれば、これらの手法はtranslation-based modelとよばれ、大きな成功を収めている。

社会課題に関連する様々なデータをナレッジグラフ化して複合的な観点からの課題解決を促進するとともに、ベクトル埋め込みを行うtranslation-based modelに基づいた欠損推定手法を適用することで、先に述べた現在の課題であるデータ形式の不一致や欠損値の推定を行い、社会課題解決を促進できると考えられる。

よって本研究では、ナレッジグラフとベクトル埋め込み型の推論手法による社会課題解決への助力を試みる。本研究では一例として、毎日多くの人間が出入りし、学外と同様に様々な社会課題が存在する、国立大学法人電気通信大学の調布キャンパス(東京都)を考える。また当大学学生課の元で運営され、学内環境改善に向けた活動を行う学生団体からデータを収集し、これを利活用して課題解決を補助する。

ここで、キャンパス内の社会課題は窃盗、落書きなど多岐にわたるが、特に狭いキャンパス内の移動手段として約6割の学生が利用する自転車に関しては、Fig. 1のように大小合わせて30箇所以上の駐輪場があり1200台以上駐輪可

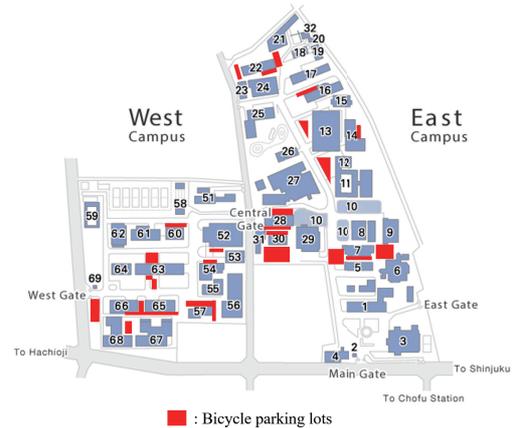


Fig. 1. Bicycle parking lots in the university

能にも関わらず、曜日や時間帯によって特定の駐輪場だけが利用率130%を超え⁽¹⁰⁾、災害時の避難経路寸断の危険性や、景観上の問題が指摘されるなど、多くの課題を抱えている。他にも、駐輪環境のマナー調査に対して4割以上が不満だと回答する⁽¹⁰⁾など、駐輪環境問題は非常に改善の余地が大きい。また、羽鳥ら⁽¹¹⁾によれば、放置駐輪は深刻な社会問題であり、全国の多くの自治体において行政上の喫緊の課題として位置づけられている。加えて国土交通省⁽¹²⁾により、駐輪場の整備が放置駐輪数の減少につながることも分かっており、駐輪環境改善は最優先事項の一つであると言える。

先述の学生団体は、これらの課題解決に向けて駐輪状況監視や、台数調査などを行っている。また、混雑している駐輪場から、空いている駐輪場へ自転車を移動させたり、利用率の低い駐輪場の情報を掲示するなどの対策も行っている。しかし、対象地域の広さに対して人手が不足しており、キャンパス全域に点在する駐輪場全ての情報を網羅的に把握することができていないため、効果は少ない。またこのように未観測な駐輪場が存在しているため、先に述べた社会での課題点と同様に、台数調査によって得た台数データ中に未観測値や欠損データが存在しているため容易に利用できない。加えて、台数データ単体では利活用が困難な上、これも社会での課題と同様に、大学が公開している他のデータは形式にばらつきがあるため連携して利活用することが難しい。よって全域での混雑状況を把握することは困難であり、対応が遅れが生じている。未観測駐輪場を含む全駐輪場の任意の日時における利用率などの情報が、なるべく短時間に少しでも高精度で推定できれば、環境改善に大いに役立てることができる。

以上より、本研究ではキャンパス内の駐輪環境問題解決に向け、ナレッジグラフとベクトル埋め込み型の欠損推定を用いたデータの利活用について考える。駐輪環境問題解決に必要な事象は、現状把握、台数増加原因の把握、解決策立案・実行の3点に分解できる。本研究ではこのうち、どの駐輪場にどの時間帯にどのくらいの台数があるかといった現状把握の部分について、以下の手順で、課題解決の一

助となるような仕組みを開発する。1. 学内データは多様な形式で公開されているため、既に観測されている自転車台数データに加え、学内環境に関するあらゆるデータをナレッジグラフとして再構築する。2. 構築したナレッジグラフを、web上に公開し、誰でも利用可能な形にする。3. 構築したナレッジグラフに、欠損推定に関する提案手法を適用し、自転車台数に関する推定効率と精度を既存手法と比較して評価する。

本研究は我々の先行研究⁽¹³⁾の拡張であり、本研究の貢献は、次の通りである。上記1.により、複合的な観点から課題解決に取り組めるスキーマを構築・提案し、学内の様々なデータの連携を容易にする。2.により、誰でも学内の諸問題解決のための情報サービスを開発可能な形にする。さらに3.によって未観測駐輪場における現状把握を効率化・高精度化し、学生団体による駐輪環境問題解決の一助とする。同時に、同様の手法を社会課題に拡大して適応する方法についても考察する。

以下、本論文では、まず2章で関連研究について述べる。次に3章でナレッジグラフ構築および欠損推定手法の改変について述べる。4章で欠損推定について既存手法と提案手法の精度比較および評価を行い、最後に5章でまとめと今後の展望を述べる。

2. 関連研究

(2・1) ナレッジグラフを用いた社会課題解決の試み

都市データに基づいて Linked Data を構築する研究はいくつかある。Lopezら⁽¹⁴⁾は、Linked Dataとしてセンサーデータを公開するためのプラットフォームを開発した。プラットフォームはセンサーデータを収集し、RDFデータとしてリアルタイムで公開する。この研究の最大の利点は、都市の複数のデータを統合および処理できることである。しかし、このアプローチは実センサの設置によるデータ取得や複数の有償ライセンスの導入によって実現されるものであり、導入および運営の維持にコストがかかるため、本研究で想定した非営利の運営主体の場合、同様の取り組みを行うことは難しい。また、Bischofら⁽¹⁵⁾は、都市に関する多数のデータが依然として手動で収集されており、相互に繋がることが困難であると指摘した上で、都市データを Linked Data に変換し、統計的回帰法と主成分分析を組み合わせることで欠損推定を行い、再配布する仕組みを提案した。しかしこの研究は任意のデータセットが複数の異なるデータソース中に存在する場合有用だが、本研究のようにデータセットが他のデータソース中に存在しない場合には適用が難しい。また、江上ら⁽¹⁶⁾は、社会課題の一つとして駅前の放置自転車問題を取り上げ、これをデータの利活用によって解決する試みを行っている。この研究では放置自転車問題に関する統一的な LOD スキーマを設計し、Social Networking Services (SNS) や行政の Web サイトから実際のデータを収集した上で、バイジャンネットワークを用いて欠損値を推定し LOD として統合している。またこれを用いた可視化

を通して、市民の問題意識向上に寄与する仕組みを提案している。但し、Bischofらや江上らは、構築したナレッジグラフに対し、一般的な推定手法を用いており、ナレッジグラフの意味的な繋がりを十分に生かした推定手法の適用を試みていない。本研究では、江上らの研究内で扱われているスキーマ設計の方法論やオントロジー等を踏襲しつつ、ナレッジグラフの強みであるデータ同士の繋がりを活かした欠損推定手法を用いて、台数推定を行い、その結果を受けて欠損推定手法の更なる効率化・高精度化を図った。

(2・2) 欠損推定手法 Bordesら⁽⁸⁾は、Wordnet⁽¹⁷⁾や Freebase⁽¹⁸⁾等の大規模ナレッジグラフ内の欠損推定において、ナレッジグラフ内のエンティティとそれらに関係づけるリレーションを多次元のベクトル空間内に埋め込み、スカラーエネルギーによって変数間の依存関係を把握する考え方を用いて類似度を算出する TransE を提案している。この研究では、トリプル (h, r, t) が与えられたとき、ベクトル空間でのスカラーエネルギーに基づいて、このトリプルのスコアは式1で表される非類似性距離 $d(h+r, t)$ に等しいと考える。ここで、トリプルとはエンティティ $h, t \in E$ (エンティティ集合) とリレーション $r \in R$ (リレーション集合) からなる主語、述語、目的語のセマンティクスを持つデータのつながりである。

$$d(h+r, t) = \|h+r-t\|_{\ell_{1/2}} \dots \dots \dots (1)$$

$\ell_{1/2}$ は、 L^1 ノルムまたは L^2 ノルムを表す。また、これを用いてトリプル (h, r, t) が正しい組み合わせの時はスコアが小さく、そうでない時はスコアが大きくなるようにベクトル埋め込みを行うために、以下の様な損失関数 \mathcal{L} を用いる。

$$\mathcal{L} = \sum_{(h,r,t) \in S} \sum_{(h',r',t') \in S'_{(h,r,t)}} [\gamma + d(h+r, t) - d(h'+r', t')]_+ \dots \dots \dots (2)$$

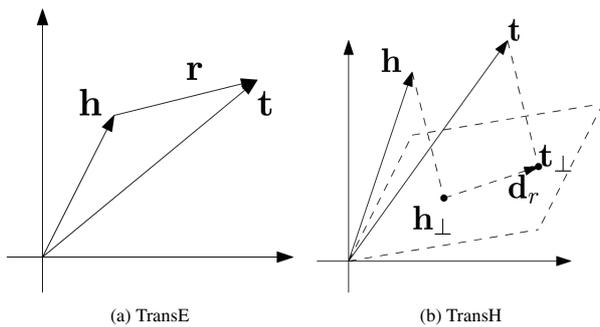
ここで、 $[x]_+$ は、 x の正成分、 $\gamma > 0$ はマージンハイパーパラメータである。また S が正例トリプル (h, r, t) の集合を表すのに対し、 $S'_{(h,r,t)}$ はトリプル中の主語部または目的語部を異なるエンティティに置換することで生成される負例トリプルの集合であり、 $h \neq h', t \neq t'$ であるとき、以下の式で表される。

$$S'_{(h,r,t)} = \{(h', r, t) | h' \in E\} \cup \{(h, r, t') | t' \in E\} \dots \dots \dots (3)$$

前述のように正例である $d(h+r, t)$ は小さく、負例である $d(h'+r', t')$ は大きくなるようにモデルを生成するために式2を最小化するようなベクトルを確率的勾配降下法を用いて学習する。式3のように片側のエンティティが欠けている欠損トリプルに対し、欠損エンティティを全てのエンティティに置換して、学習の結果生成されたモデルによるスコアリングを行い、値が小さいほど正解に近いという結果が得られる。これにより、ナレッジグラフ内の欠損を補完することが可能になる。この手法の表現性をさらに拡張したものととして Fig. 2 のような Wangら⁽¹⁹⁾の提案手法 transH

Table 1. Data and citation sources

Data	Citation Source
Time, e.g., fiscal year, semester, month, day, date, time zone, etc.	Sources listed below
Parking areas and the number of bicycles	Student group materials (not open to the public)
Course titles, and classrooms	Education Affairs Division web page ¹
Seating capacity in every room	Education Affairs Division materials (not open to the public)
Names of rooms in every building	Facilities section materials (not open to the public)
Event titles and venues	University website event calendar ²
Temperature and precipitation	Japan Meteorological Agency website ³
Latitude and longitude of every place	Google Maps ⁴

¹ <http://kyoumu.office.uec.ac.jp/timet/>² <https://www.uec.ac.jp/event/>³ <https://www.data.jma.go.jp/obd/stats/etrn/select/prefecture00>⁴ <https://www.google.co.jp/maps>Fig. 2. Overview of TransE and TransH⁽¹⁹⁾

がある。この手法では、Bordes らの TransE におけるベクトル埋め込みの手法を拡張し、リレーション固有の超平面上へエンティティを写像することで異なるリレーションを持つエンティティ同士の表現性を増幅させ、精度向上に貢献している。具体的には、TransE 中のエンティティ h, t を式 4 でリレーションごとに異なる超平面上に写像し、 h_{\perp}, t_{\perp} とする。

$$h_{\perp} = h - w_r^T h w_r, \quad t_{\perp} = t - w_r^T t w_r \dots \dots \dots (4)$$

超平面上のリレーション d_r を用いて、スコアリング関数は TransE と同様に以下のようにになる。

$$d(h_{\perp}, r, t_{\perp}) = \|(h - w_r^T h w_r) + d_r - (t - w_r^T t w_r)\|_2^2 \dots (5)$$

さらにこの他にも、transE の表現の拡張として、リレーションごとに異なる空間上にエンティティを写像する Lin ら⁽²⁰⁾ の transR や、トリプルのうち、主語部と目的語部で写像ベクトルを分ける Guoliang ら⁽²¹⁾ の transD などがある。

しかしながらこれらの研究は、様々な分野や問題にまたがるあらゆる情報を格納した大規模なナレッジグラフに関しての推定を基本としており、学内の様な小規模なドメインで台数推定などの身近なトピックを扱う場合には効果が保証されていない。したがって本研究では、これらの手法が身近なデータセットを用いた小規模なナレッジグラフ上でどの程度の推定精度を達成できるか検証し、検証結果から独自の改変を加えることで推定効率と精度の向上を試みた。

Table 2. Prefix and URI for each name space

Prefix	URI
aiiso:	http://purl.org/vocab/aiiso/schema#
event:	http://purl.org/NET/c4dm/event.owl#
geo:	http://www.w3.org/2003/01/geo/wgs84_pos#
gn:	http://www.geonames.org/ontology#
ipb:	http://www.ohsuga.is.uec.ac.jp/bicycle/resource
ipbv:	http://www.ohsuga.is.uec.ac.jp/ipblod/vocabulary#
owl:	http://www.w3.org/2002/07/owl#
rdf:	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs:	http://www.w3.org/2000/01/rdf-schema#
time:	http://www.w3.org/2006/time#
teach:	http://linkedscience.org/teach/ns#
wo:	http://www.auto.tuwien.ac.at/downloads/thinkhome/ontology/WeatherOntology.owl

3. 学内ナレッジグラフ生成と欠損推定

3.1 学内ナレッジグラフ生成

3.1.1 データセット まず、学生スタッフが保有する、学内の主要駐輪場を含む各駐輪場での自転車台数データをはじめ、本学 HP、Google Maps、気象庁 Web サイトなどから、Table 1 に示す項目の 2017 年度のデータを抽出した。

3.1.2 スキーマ設計 次に、これらを RDF として構造化するために Fig. 3 のようなスキーマを独自に設計した。これにより、一意なエンティティ同士を繋いで複数のデータが関連性を持つ構造を実現した。今回扱っている駐輪環境問題について、本研究では、学内で開催されるイベントや授業と同様、一つのイベントとして捉えられると考え、同様の考え方で都市部における放置自転車問題を解決するために設計された放置自転車 LOD⁽¹⁶⁾ を再利用している。これに加え、各データの再利用性を高めるため、オントロジーを新規に構築せず、Table 2 に示す 12 オントロジー、および各オントロジーに属する 25 プロパティ、10 クラスを用いてスキーマを構築した。

これによって、オントロジーの制約に基づく推論が可能になると同時に、データ構造を理解する手間が省けることで、他のスキーマとリンクさせて二次利用する際の利便性などを向上させられるメリットがある。また実用性に関しても、自転車に関するエンティティを中心として、ナレッ

Bicycle parking lot URI	Surrounding population	Bicycle parking lot URI	The number of bicycles	The number of lectures held in the morning
http://www.ohsuga.lab.uec.ac.jp/_bike_num_2017_4_100_pA	669	http://www.ohsuga.lab.uec.ac.jp/_bike_num_2017_4_100_pA	169	17
http://www.ohsuga.lab.uec.ac.jp/_bike_num_2017_4_100_pB	812	http://www.ohsuga.lab.uec.ac.jp/_bike_num_2017_4_100_pB	83	19
http://www.ohsuga.lab.uec.ac.jp/_bike_num_2017_4_100_pCOB	82	http://www.ohsuga.lab.uec.ac.jp/_bike_num_2017_4_100_pCOB	-	2
http://www.ohsuga.lab.uec.ac.jp/_bike_num_2017_4_100_pCOF	294	http://www.ohsuga.lab.uec.ac.jp/_bike_num_2017_4_100_pCOF	0	3
http://www.ohsuga.lab.uec.ac.jp/_bike_num_2017_4_100_pE2NA	82	http://www.ohsuga.lab.uec.ac.jp/_bike_num_2017_4_100_pE2NA	56	2
http://www.ohsuga.lab.uec.ac.jp/_bike_num_2017_4_100_pBE2NI	82	http://www.ohsuga.lab.uec.ac.jp/_bike_num_2017_4_100_pBE2NI	1	82

Fig. 5. Example of data extracted using SPARQL query

ティを辿って開講授業エンティティにリンクできているからである。具体的には、まず任意の時間的値および場所を持つ駐輪場データから駐輪場の周辺施設を抽出する。次に、その施設に所属する教室を抽出し、その教室で実施されている授業を抽出する。さらに、それらの授業のうち、指定された時間的値と繋がりを持ち、かつ時限が午前中を表す1限または2限である授業のみを抽出する。これを集計することで、任意の駐輪場データから、午前中の開講授業数を取得している。この抽出例のように、提案したスキーマは、駐輪場のインスタンスを起点として全てのノードから間接的にリンクされている。また、このようにナレッジグラフを用いてデータ連携を図ることで、様々なデータを関連付けて抽出することが可能となり、駐輪場の設置場所の選定や、時間割設計の効率化、学生の動線を考慮した教室選択などが容易になると考えられる。

〈3・2〉 欠損推定 収集したデータには、未観測であるなどの理由で一部の情報のみ抜け落ちた欠損が存在している。特にこのような欠損は、自転車台数のリテラル値を目的語とするトリプル(以下、台数トリプル)に多く見られる。これは、1章で述べたように人手と時間の不足によって、学生団体が全域の駐輪場の台数を確認することが困難なためである。このためこれらのデータを集約しているナレッジグラフ中にも、〈駐輪場ノード, rdf:value(プロパティ), 欠損値〉の様な形で、多くの欠損値が存在する。事例として、データ抽出結果である Fig. 5 中、右図の上から3行目の台数が空白となっていることなどが挙げられる。本研究では1章で述べた通り自転車台数データ中に多くの未観測データが存在することから、先ほど構築したナレッジグラフに対して、知識の繋がりを利用して台数推定を行う。

〈3・2・1〉 既存手法による欠損推定 まず、TransE⁽⁸⁾, TransH⁽¹⁹⁾, TransR⁽²⁰⁾, TransD⁽²¹⁾の既存4手法を実装して、台数推定を行った。これらの手法は、ナレッジグラフからセマンティクスを持ったデータの繋がりであるトリプルを全て抽出し、トリプルの構成要素であるエンティティとリレーションをベクトル空間へ埋め込み、欠損を推定する。先述のように、上記4手法を含めた既存のベクトル埋め込み及び欠損推定手法をまとめて translation-based モデルと呼ぶが、これらのモデルでは、評価のためにモデル内で新たな欠損を生成しこれを補完するという方法をとるため、本研究では総トリプル 20889 のうち、データ格納時点で既に欠損のあった 372 トリプルを除いた 20517 トリプルを実験

Table 3. Estimation accuracy of existing method with campus knowledge graph

	MeanRank	Hits@10	Hits@3	Hits@1
TransE	89.747	0.351	0.284	0.208
TransH	70.479	0.390	0.274	0.219
TransR	289.414	0.205	0.086	0.023
TransD	148.131	0.288	0.208	0.051
Random	1298	0.385	0.116	0.039

に用いた。そのうち、台数トリプルを含む 19868 トリプルを訓練データとした。

実験の条件としては、[学習率, 隠れ層, ミニバッチサイズ, マージンハイパーパラメータ]=[0.001, 100, 100, 1.0]として 1000 エポックの訓練とモデルの評価を各手法 5 回ずつ行なった。なお、本研究では学内自転車の台数に関する欠損を推定することを目的としているため、モデルの評価には、前述の 20517 トリプルのうち、訓練データに含まれず、自転車台数のリテラル値を目的語とする 649 の台数トリプル抽出し、テストデータとして用いている。この実験の結果は、Table 3 の通りである。

なお、Table 3 中の MeanRank は、translation-based モデル中におけるベクトル空間でのスカラーエネルギーに基づいた距離計算 $d(h, r, t)$ によって算出された各トリプルでのスコアの平均であり、値が低いほどより精度が高いといえる。一方、Hits@n (n=10, 3, 1) は、全トリプルにおける実験中で、正解エンティティが上位 n 位までに予測された割合を表しており、値が高いほど精度が高いといえる。ここで、最下段はランダムに台数を推定した場合の数値となるが、精度を特に精密に表す Hits@3 と Hits@1 の 2 指標において、TransR を除く全ての既存手法がランダムの場合を上回っており、ナレッジグラフの意味的關係性が台数推定に一定の効果を及ぼしていることが分かる。4 手法中、ほぼ全ての評価指標において TransH の精度が最も高いが、これは当該手法が主語部と目的語部で 1 対多または多対 1 の関係にあるトリプルに優位に推定を行えるという特徴を持っているからだと推測できる。今回扱っている台数推定では、〈駐輪場ノード, rdf:value (プロパティ), 台数 [リテラル値]〉といった 3 ノードで台数トリプルが構成されており、主語部と目的語部の関係はほぼ全てのプロパティにおいて多対 1 となる。したがって、TransH による結果が最も高い精度であったと考えられる。

〈3・2・2〉 既存手法における課題 〈3・2・1〉項および Table 3 より, 既存手法が台数推定に効果的であり, 中でも TransH が今回の台数推定において最も理想的な手法であることが分かった。しかし 1000 エポックの訓練を 5 回行った結果である Table 3 中, TransH における Hits@10 を見ても, 正解台数が上位 10 位以内に推定される確率は 0.390 と低く, 学生団体がリアルタイムで全域の駐輪状況を把握し対策をとるのに用いる結果としては, まだ十分でないといえる。

〈3・2・3〉 提案手法 本研究では, ナレッジグラフの強みは, リレーショナルデータベース等では表現性に欠けるデータ間の関数従属性を容易に表現できる点だと考え, 台数推定を行う既存研究にこの考え方を反映した手法を提案する。具体的には, 自転車台数と, その台数が観測された日時および場所における周辺の人口に関数従属性があるとみて, ナレッジグラフ中における周辺人口のデータを, より強く反映させた台数推定を行う。自転車台数が観測された日時と同条件で当該駐輪場周辺にいたと思われる人口を仮定的に算出して, この数値を TransH 中のベクトル写像およびスコアリングに反映させることでより精度の高い推定を目指した。なお, 人口の算出には, 日頃学内で最も人口変動に影響すると思われる授業カリキュラムと, 授業が開講される教室の座席数[†]を用い, 駐輪場から半径 50 m 以内の教室をこの対象とした。

提案手法では, 任意の日時と場所を持つ駐輪場エンティティ h_1 に対して, 台数との繋がりを表す台数トリプル (h_1, r_1, t_1) および, 周辺人口との繋がりを表すトリプル (h_1, r_2, t_2) (以下, 周辺人口トリプル) を考える。それぞれのトリプルに対して, TransH 中の式 4 でリレーション固有の超平面を生成し, 超平面上にエンティティを写像する。Fig. 6 および Fig. 7 は, それぞれ図中左側に表す台数トリプル (h_1, r_1, t_1) および周辺人口トリプル (h_1, r_2, t_2) を, 図中右側の様に超平面上に写像する様子を表す。各トリプルの写像後をそれぞれ, $(h_{1\perp}, d_{r_1}, t_{1\perp})$, $(h'_{1\perp}, d_{r_2}, t_{2\perp})$ とする。写像後のトリプルに含まれる各エンティティは, 式 6 および式 7 のように算出される。ここで, w_{r_1} および w_{r_2} は, 単位ベクトルであり, リレーション r_1 および r_2 固有の超平面を表す。また, d_{r_1} および d_{r_2} は, 各超平面上の並進ベクトルを表す。さらに, Fig. 8 に示すように, 算出された写像後のエンティティ同士の差をとり, 既存手法中の式 5 で表されるスコアリング関数に加えることで, 台数エンティティと周辺人口エンティティを関連づけてスコアを算出する。

$$h_{1\perp} = h_1 - w_{r_1}^T h_1 w_{r_1}, \quad t_{1\perp} = t_1 - w_{r_1}^T t_1 w_{r_1}, \quad \dots \quad (6)$$

$$h'_{1\perp} = h_1 - w_{r_2}^T h_1 w_{r_2}, \quad t_{2\perp} = t_2 - w_{r_2}^T t_2 w_{r_2}, \quad \dots \quad (7)$$

台数トリプルにおける目的語部 (台数) の欠損推定に周辺人口トリプルの写像式 7 のうち主語部の式を反映させるこ

[†] 本学教務課資料 (非公開) より取得, 未掲載の教室に関しては, 本学施設課資料 (非公開) より床面積を参考に算出。

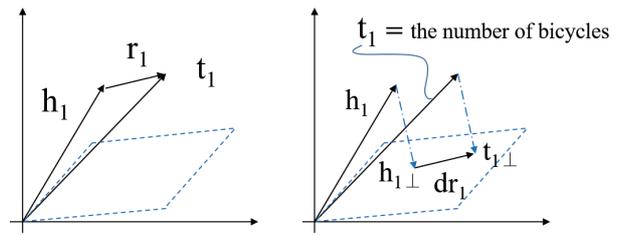


Fig. 6. Translating the relationships of the number of bicycles on hyperplane

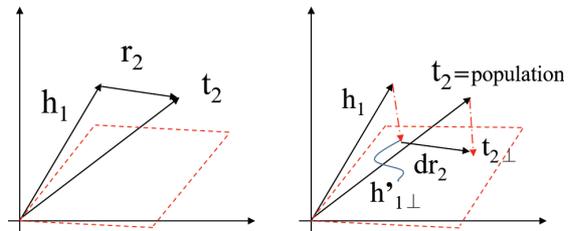


Fig. 7. Translating the relationships of the population on hyperplane

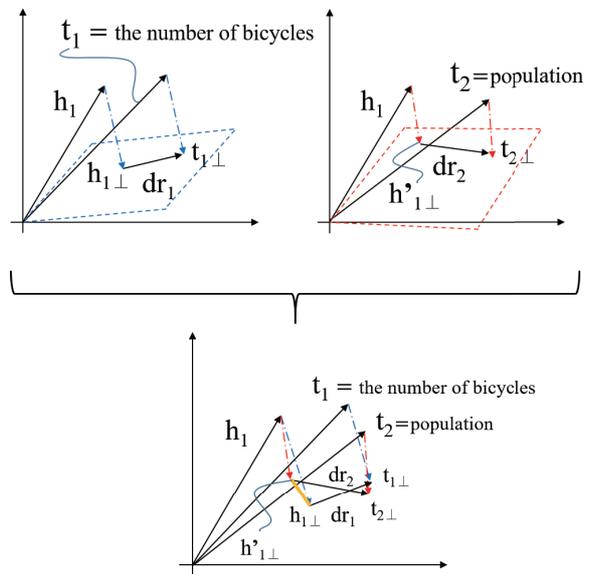


Fig. 8. Overview of the proposed method

とで, スコアリング関数は, 既存手法における式 5 に対し, 式 8 の様に表される。

$$d(h_{1\perp}, r_1, t_{1\perp}) = \|(h_1 - w_{r_1}^T h_1 w_{r_1}) + d_{r_1} - (t_1 - w_{r_1}^T t_1 w_{r_1})\|_2^2 + \|(h_{1\perp} - h'_{1\perp})\|_2^2 \dots \dots \dots (8)$$

4. 評価

〈4・1〉 評価方法の修正 ここで, 各既存手法に用いられている前述の評価指標について考えると, 特に Hits@n に関しては, 正解とされたエンティティをどれだけの高順位で推定できるか, という順位の上位層だけをみた評価手法となっている。これは, 〈2・2〉節で述べた様に, 既存手法で扱っている知識ベースは大規模なナレッジグラフであり,

Table 4. Number estimation accuracy of the existing method and the proposed method

Method	Sum of the absolute value error						
	Minimum	Maximum	1,000 epoch	5,000 epoch	10,000 epoch	15,000 epoch	20,000 epoch
Existing method	37237	37430	37325	37269.8	37279	37289.6	37302.2
Proposed method	37239	37285	37283	37249.8	37245.7	37247.33	37247.7
Evaluation value difference	-2	145	42	20	33.3	42.27	54.5

※ The difference between the evaluation values is positive when the proposed method outperforms the existing one.

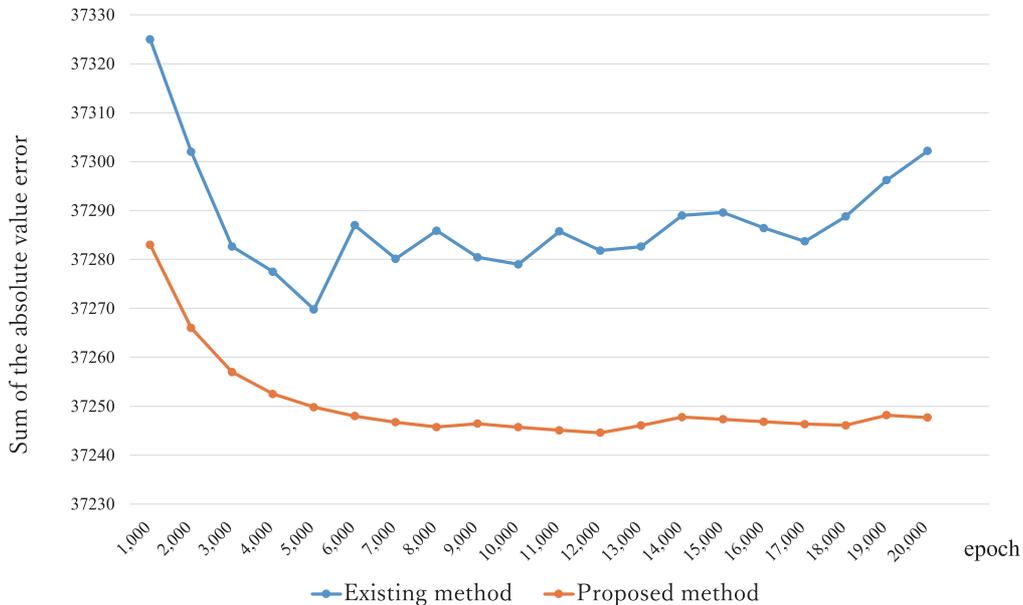


Fig. 9. Transition of number estimation accuracy by learning of existing method and proposed method

幅広く広がる種々の分野から、様々なエンティティが推定結果の選択肢として選ばれるためである。よってスコアリング関数内で置換されるエンティティの種類も多岐にわたる上、既存手法中では主に数値推定ではなくノード推定を行った場合の精度を評価するため、生成されたモデルの精度が平均してどの程度であるかといった評価は難しく、順位全体の全体像を見た評価が行えていない。

一方で、本研究で扱う台数推定においては、テストデータとして台数トリプルのみを与えるため、スコアリング関数内で置換される全てのエンティティが数値を持っており、モデルの順位予測が平均してどの程度の値を推定しているかという評価が可能である。具体的には、1000エポックの訓練によって生成されたモデルによる、台数の推定順位全体を見てその加重平均を算出し、これと正解台数との絶対値誤差をモデルの評価値とする方法である。加重平均は、各変量の“重み”を加味した平均のことであり、式9のように求められる。

変量： $x_1, x_2, x_3 \dots, x_n$ に対する変量の加重（重み）： $w_1, w_2, w_3 \dots, w_n$ とするとき、加重平均を \bar{x} とすると

$$\bar{x} = \frac{\sum_{i=1}^n x_i w_i}{\sum_{i=1}^n w_i} \dots \dots \dots (9)$$

今回の台数推定では重みづけの明確な基準が存在しないため、最も確からしいと推定された台数に重みとして全台数

トリプル数である 649 を与え、そこから順位が 1 下がると重みを 1 下げ、最下位で推定された台数には重み 1 が与えられるように重みづけを行なった。このような設定とした理由を以降に述べる。簡単のため台数データとして、項数 600 初項 1 公差 1 の等差数列を考えると、重みの加減を m としたとき、 $m \geq 0.5$ の場合、加重平均の取りうる値の範囲にはほぼ変化が現れない。例えば m を、今回与えている $m = 1$ より極端に大きい $m = 100$ とした場合でも、 $m = 1$ の場合との差は、最小値で 0.2% 未満、最大値では 0.1% 未満であり、自然数以上の議論が重要になる台数においてはこの差は無視できる範囲である。またこのように m を大きく与える場合、上位の値を積極的に評価する従来手法と同様の評価になってしまうため、新たに評価手法を設定する意味を為さない。逆に m を非常に小さくしていくと、徐々に取りうる値の範囲が狭まり、最終的に最大値および最小値が同じ値に収束してしまうため、これもまた評価としての意味を為さない。よって本研究では、妥当な値として $m = 1$ を設定している。また提案した評価手法の範囲内では、加重平均の取りうる値の範囲には、有用性に影響するほどの大きな変化は生じない。

この方法で算出された加重平均と正解台数との絶対値誤差をとり、生成されたモデルの評価とした。この方法で評価した場合、モデルの精度が高いほど、絶対値誤差は小さ

くなる，つまり評価値が小さいほど精度が高いといえる。

〈4・2〉 既存手法と提案手法の比較 本研究内で扱う既存手法および提案手法の実装にあたっては，訓練トリプルの選定に擬似乱数生成式である線形合同法を用いている。線形合同法は0から総訓練トリプル数である $M = 19868$ の間の整数の乱数を生成し，最大周期 M 中に同じ値は2度出力されないという特徴を持っているため，今回は1度の訓練を1000エポック，つまり1度の訓練につき学習するトリプル数を1000トリプルとし，これを20回繰り返すことで，総訓練トリプル数に極めて近い20,000トリプルの学習を行ない，〈4・1〉節で提案した評価方法に基づいて精度の確認を行なった。なお，本論文では，複数通り行なった実験中で最も精度の高かった場合について既存手法と提案手法の精度を比較する。

既存手法である TransH および，改変を行なった提案手法で20,000エポックの訓練と推定を繰り返し，提案した評価方法を適用した結果，Table 4 および Fig. 9 の様になった。

〈4・3〉 考察 Table 4 より，既存手法と提案手法で最小値には大きな変化がないが，最大値を大きく抑え，20,000エポックの実験の平均値においても，評価値の差で54台以上正しく推定できていることが分かった。649件分の駐輪場データを扱っていることを考えると，任意の日時および場所における1駐輪場あたりの推定精度は平均して約0.084台上昇していると考えられる。

また，Fig. 9 は学習過程における精度の推移を表しているが，提案手法は一貫して既存手法の誤差を下回っていることが分かる。さらに，5,000エポック以降で既存手法に過学習の傾向が見られるのに対して，提案手法では20,000エポック終了時点でも一定以下の誤差を維持しており，より汎用的なモデルを生成できていると言える。次に，既存手法が最も高精度となる5,000エポックまでの動きに着目すると，提案手法は2,000エポックの訓練を終了した時点で既存手法の5,000エポック終了時の精度を超えており，同等の精度に到達するのが既存手法と比較して62%早くなっていることが分かる。これにより，突然のイベント開催や教室変更などに対する台数予測も，既存手法より効率的に行うことができると考えられる。

以上より提案手法は，既存手法比較して，より高効率かつ高精度に台数推定を行うことができた。

なお予備実験として，観測日の1時限目の台数のみを全時間帯の駐輪データに反映させる実験も行なった。これは，駐輪場利用者は1時限目の授業時に自転車を駐輪後，授業間の移動では自転車を移動させない場合が多いとの仮定に基づくもので，結果として既存手法を上回る精度での台数推定には成功したが，Table 4 および Fig. 9 に示した提案手法には劣る結果となった。以上より，各時間帯における，観測駐輪場周辺の授業の開講数およびこの授業に出席するであろう人口が，駐輪台数に影響を及ぼしているという事実が明らかになった。また同時に，この事実を台数推定に反映させる提案手法において，式8の試みが，有用であるこ

とが示された。これによって提案手法では，ナレッジグラフ中でより強く欠損推定に影響させたいノードを恣意的に選択することができ，今回扱った学内データの例のように，データ中で明示的な関数従属性が見られる場合に，関係性をより強調して学習させることが可能になったと言える。

また，現時点ではまだ検証できていないが，この欠損推定手法は規模のドメイン的に制約を受けないため，いかなるナレッジグラフにも適応可能で，大学内のみならず，広範な社会課題等のナレッジグラフにおける数値推定においても活用できると考えられる。1章で述べた様に，社会課題とその解決について内閣府⁽⁹⁾は，オープンデータ化やデータ連携などのデジタル技術による対応を急いでいる。今回提案した手法を用いて様々なデータを連携し，それを活用することによる台数推定またこれを応用させた人口推定などを通じて，例えば都市部や駅周辺など，より広範囲なドメインで，多くの人手や時間を必要とする自転車台数把握をはじめ，これを用いた放置駐輪問題の是正など，より規模の大きな社会課題の解決に適用可能だと考えられる。また，今回扱ったスキーマは大学内のデータを有効に活用するために独自に設計したものであり，提案手法では大学内で問題とされている駐輪環境問題の解決に向けて，自転車台数データの欠損推定を行ったが，当スキーマ上の他ノードの推定や，本研究のように様々な情報が融合したほかのナレッジグラフのスキーマにおいても今回と同様の方法で活用可能だと考えられる。少ない訓練回数で台数推定精度が向上する提案手法の効率性と精度を生かし，突然の自然災害や状況の変化に対応した短時間での予測や，数千万から数十億トリプルを持つような大規模なナレッジグラフにおける推定にも応用できる可能性があり，社会課題解決に向けて更なる貢献が期待できる。

5. おわりに

〈5・1〉 本論文のまとめ 本論文では，内閣府の最新の動きなどを踏まえて，複数のデータを連携させることで，従来とは異なる複合的な視点からの社会課題解決につながる可能性を示唆した。その上で最も身近な大学内の環境を一例として考え，特に学内で問題視されている駐輪環境問題において，社会と同様にデータ形式の不一致や欠損値の存在などの課題があることを確認した後，ナレッジグラフへの統合と欠損推定を行なった。具体的には，データの再利用性を確保した上で学内情報に特化した独自のスキーマを設計し，様々な形式でWeb上に散らばっていたデータを収集してこのスキーマ上で独自のナレッジグラフとして統合した。また，このナレッジグラフに欠損推定で広く知られるナレッジグラフ特有の既存手法を適用し，台数推定の精度を確認後，一部の欠損推定手法を改変することで台数推定効率と精度の向上を達成した。

〈5・2〉 今後の展望 今後の展望として，次の3点を考えている。1点目に，混雑状況の可視化などのサービス化である。現状の台数推定効率および精度を鑑みると，サー

ビス化につなげるには時間を要するが、今後の展望として他にも、駐輪場推薦、避難誘導をはじめ、駐輪場設置場所提案、開講場所効率化などの各種サービス化を行いたい。身近な題材であることから、これらのサービス化によって、キャンパスを利用する教職員や学生をはじめ、学内管理を委託されている学生団体や、大学の運営主体など幅広いセグメントのサービス利用者が、利益を享受可能であると考えている。2点目に、都市部や駅周辺など、より開かれたドメインにおける提案手法適用である。本研究では大学キャンパス内という閉鎖的なドメインでのデータ収集および活用を行った。今後は、より広範囲なドメインにおいてデータを収集し、ナレッジグラフ構築を行いデータ活用基盤を生成するとともに、提案手法による欠損推定を行い、自転車台数把握や放置駐輪の是正などを行いたいと考えている。3点目に、さらに精度を向上させる、より良いアルゴリズムの提案である。4章で述べたように、提案した手法によって、特に少ない訓練回数での台数推定精度は向上したものの、現状のアルゴリズムではいまだに有用な推定精度を得られているとは考えにくい。1章で述べたように、本学では駐輪場によって利用率が130%を超えることがあり、これが災害時の危険性や景観上の問題として指摘されている。学生団体は、これらの課題を回避するために駐輪台数を把握し、対策実施の是非を判断する。よって、危険性等が問題視される130%の駐輪台数を混雑台数と定義すると、満車に達していない状態(100%以下)と混雑台数以上の状態(130%以上)を判別できる、すなわち誤差が駐輪可能台数の30%以内となることが実用化において重要になる。桑原ら⁽¹⁰⁾より、特に混雑する駐輪場の駐輪可能台数(満車時)が60台に設定されていることを考えると、現在全649トリプルで生じている37249.7の誤差は、一つの駐輪場データあたり約57.4台であり、駐輪可能台数60台の駐輪場で誤差を30%に留めるためには、これを18台程度まで改善させる必要がある。よってさらなるアルゴリズムの改良が必要だと考える。

以上の展望により、ナレッジグラフおよび欠損推定を用いたデータ利活用による社会課題解決を促進したい。

謝 辞

本研究は、JSPS 科研費 JP17H04705, JP18H03229, JP18H03340, JP18K19835, JP19H04113, JP19K12107 及び JST, さきがけ, JPMJPR1934 の助成を受けたものです。本研究を遂行するにあたり、研究の機会と議論・研鑽の場を提供して頂き、御指導頂いた早稲田大学 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へ感謝致します。またご多忙中、日頃から多くのご指摘ご助言をくださいました、協力研究員の江上周作氏に、この場を借りて心から感謝の意を表します。

文 献

- (1) 本計画, <https://cio.go.jp/node/2413> (2019/10/15 参照)
- (2) 総務省: 地方公共団体のオープンデータの推進, http://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/opendata/ (2019/10/15 参照)
- (3) 内閣府: Society5.0 実現に向けたデータ連携基盤現状と課題, <https://www8.cao.go.jp/cstp/tyousakai/datarenkei/1kai/siry03.pdf> (2019/10/15 参照)
- (4) 内閣府: データ連携基盤の技術的検討事項, <http://www8.cao.go.jp/cstp/tyousakai/datarenkei/2kai/siry01.pdf> (2019/10/15 参照)
- (5) ontotext: *What are Linked Data and Linked Open Data*, <https://www.ontotext.com/knowledgehub/fundamentals/linked-data-linked-open-data/> (2019/10/15 参照)
- (6) L. Ehlinger and W. Wöb: "Towards a Definition of Knowledge Graphs", Proc. of the 12th International Conference on Semantic Systems (SEMANTICS) (2016)
- (7) H. Yoon, H. Song, and S. Park: "A Translation-based Knowledge Graph Embedding Preserving Logical Property of Relations", Proc. of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT), pp.907-916 (2016)
- (8) A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and L. Yakhnenko: "Translating Embeddings for Modeling Multi-relational Data", Proc. of Neural Information Processing Systems 26 (NIPS), pp.2787-2795 (2013)
- (9) H. Mousselly-Sergieh, T. Botschen, I. Gurevych, and S. Roth: "A Multimodal Translation-Based Approach for Knowledge Graph Representation Learning", Proc. of the 7th Joint Conference on Lexical and Computational Semantics, pp.225-234 (2018)
- (10) 桑原 周・金井秀明: 「監視情報の告知がもたらす放置駐輪の抑制による規範意識調査」, 情報処理学会研究報告 (2015)
- (11) T. Hatori, S. Mikiya, and S. Fujii: "An Examination on the Effects of Measures for Reducing the Illegal Bicycle Parking by Psychological Strategy: A Case Study On Toritsu-daigaku Station, Tokyu Corporation", Infrastructure planning review, pp.797-805 (2009) (in Japanese)
- 羽鳥剛史・三木谷智・藤井 聡: 「心理的方略による放置駐輪削減施策の効果検証: 東急電鉄東横線都立大学駅における実施事例」, 土木計画学研究・論文集, pp.797-805 (2009)
- (12) 国土交通省都市局: 自転車等駐車場の整備のあり方に関するガイドライン, <https://www.mlit.go.jp/common/000231328.pdf> (2019/10/15 参照)
- (13) Y. Tsukagoshi, T. Kawamura, Y. Sei, Y. Tahara, and A. Ohsuga: "Knowledge Graph of University Campus Issues and Application of Completion Methods", Proc. of the 21st International Conference on Information Integration and Web-based Applications & Services (iiWAS) (2019)
- (14) V. Lopez, S. Kotoulas, M. L. Sbodio, M. Stephenson, A. Gkoulalas-Divanis, and P. M. Aonghusa: "QuerioCity: A Linked Data Platform for Urban Information Management", Proc. of the 11th International Semantic Web Conference (ISWC), pp.148-163 (2012)
- (15) S. Bischof, C. Martin, A. Polleres, and P. Schneider: "Collecting, Integrating, Enriching and Republishing Open City Data as Linked Data", Proc. of the 14th International Semantic Web Conference (ISWC), pp.57-75 (2015)
- (16) S. Egami, T. Kawamura, Y. Sei, Y. Tahara, and A. Ohsuga: "Proposal of Eco-Cycle for Solving Illegally Parked Bicycles using Linked Open Data", journal of the Japanese Society for Artificial Intelligence, AI30-K (2016) (in Japanese)
- 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」, 人工知能学会論文誌, AI30-K (2016)
- (17) G. Miller: "WordNet: a Lexical Database for English", Communications of the Association for Computing Machinery (ACM), Vol.38, No.11, pp.39-41 (1995)
- (18) K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor: "Freebase: a collaboratively created graph database for structuring human knowledge", Proc. of the ACM SIGMOD international conference on Management of data, pp.1247-1250 (2008)
- (19) Z. Wang, J. Zhang, J. Feng, and Z. Chen: "Knowledge Graph Embedding by Translating on Hyperplanes", Proc. of the 28th association for the Advancement of Artificial Intelligence Conference (AAAI), pp.1112-1119 (2014)
- (20) Y. Lin, J. Zhang, Z. Liu, M. Sun, Y. Liu, and X. Zhu: "Learning Entity and Relation Embeddings for Knowledge Graph Completion", Proc. of the 29th association for the Advancement of Artificial Intelligence Conference (AAAI), pp.2181-2187 (2015)
- (21) J. Guoliang, H. Shizhu, X. Liheng, L. Kang, and Z. Jun: "Knowledge Graph Embedding via Dynamic Mapping Matrix", Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL), pp.687-696 (2015)
- (22) Open Link Software: *About OpenLink Virtuoso*, <https://virtuoso.openlinksw.com> (2019/10/15 参照)

(1) 内閣府: 世界最先端デジタル国家創造宣言・官民データ活用推進基

塚越雄登 (非会員) 2019年3月電気通信大学情報理工学部総合情報学科卒業。同年同大学大学院情報理工学研究科情報学専攻入学。主としてセマンティック Web・自然言語処理を用いた社会課題解決手法の研究に従事。



川村隆浩 (非会員) 1994年早稲田大学大学院理工学研究科電気工学専攻修士課程修了。同年, (株) 東芝入社。2001-2002年米国カーネギー・メロン大学ロボット工学研究所客員研究員。2003年-2018年電気通信大学大学院情報理工学研究科客員准教授。2007年より大阪大学大学院工学研究科非常勤講師。2015年-2018年科学技術振興機構情報分析室主任調査員。2019年より特任フェロー。2018年法政大学理工学部非常勤講師。2019年より農業・食品産業技術総合研究機構本部データマネジメント統括監。博士 (工学)。2012年国際会議 ISWC 10-Year Award 受賞。2013年, 2019年人工知能学会研究会優秀賞受賞。人工知能学会理事, 代議員等を歴任。主に非構造化データからの知識抽出とナレッジグラフ構築および分析・活用に従事。



清雄一 (非会員) 2009年東京大学大学院情報理工学系研究科博士後期課程修了。同年 (株) 三菱総合研究所入社。2013より年電気通信大学。現在, 同大学大学院情報理工学研究科准教授。博士 (情報理工学)。エージェント, プライバシ保護技術等の研究に従事。2016年度土木学会水工学論文賞, 情報処理学会論文賞受賞。情報処理学会, 電子情報通信学会, 日本ソフトウェア科学会, IEEE Computer Society 各会員。



田原康之 (非会員) 1991年東京大学大学院理学系研究科数学専攻修士課程修了。同年 (株) 東芝入社。1993~1996年情報処理振興事業協会に出向。1996~1997年英国 City 大学客員研究員。1997~1998年英国 Imperial College 客員研究員。2003年国立情報学研究所着任。2008年より電気通信大学准教授。博士 (情報科学) (早稲田大学)。エージェント技術, およびソフトウェア工学などの研究に従事。情報処理学会, 日本ソフトウェア科学会会員。



大須賀昭彦 (正員) 1981年上智大学理工学部数学科卒業。同年 (株) 東芝入社。同社研究開発センター, ソフトウェア技術センター等に所属。1985~1989年 (財) 新世代コンピュータ技術開発機構 (ICOT) 出向。2007年より電気通信大学。現在, 同大学大学院情報理工学研究科教授。2017年より同大学大学院情報システム学研究科研究科長併任。2012年より国立情報学研究所客員教授兼任。工学博士 (早稲田大学)。情報処理学会フェロー。ソフトウェア工学, エージェント, 人工知能の研究に従事。1986年度及び2016年度情報処理学会論文賞, 2013年度人工知能学会研究会優秀賞, 2014年度同学会功労賞, 2018年度電子情報通信学会 ISS 活動功労賞受賞。IEEE Computer Society Japan Chapter Chair, 人工知能学会理事, 日本ソフトウェア科学会理事, 同学会監事等を歴任。情報処理学会, 電子情報通信学会, 人工知能学会, 日本ソフトウェア科学会, IEEE Computer Society 各会員。

