

## 業績

申請者はこれまでのデータ工学に関わる研究を分野にとらわれずに行ってきた。(文献のリファレンスは下記「**主要な論文リスト**」を参照されたい。)卒業研究から博士研究にかけて、ファセット検索を半構造データに応用する研究を行った [1,2,3]。その際に、[3]の論文が評価され、**情報処理学会山下記念研究賞**を受賞した。加えて、博士論文の内容が評価され、**筑波大学コンピュータサイエンス専攻長表彰**を受賞した。同時期に、ファセット検索のさらなる応用可能性を模索するために、Twitter リストの検索 [4] や顔画像検索 [5] にもチャレンジした。[5] については、筑波大学内のプロジェクトとして発起し、研究活動や成果が認められ、**筑波大学コンピュータサイエンス専攻長表彰**を受賞した。博士号取得後は、知識データの公開方法として注目を集める Linked Open Data (LOD) の活用を中心テーマに研究を行ってきた [6,7,8]。LOD は異種情報ネットワーク (HIN) の一種である。LOD に関する分析手法の提案 [6] は、国際会議 iiWAS 2015 にて **Best paper award** を受賞した。同時期に、所属していたプロジェクトの研究として、データ工学に関する研究成果 (ストリームデータ処理、可視化、対話的検索など) の社会実装を行った [9,10,11]。これらのシステムの一部は現在もフィールドである藤沢市やつくば市で稼働している。これらの主要成果が示すように、申請者はデータ工学の様々な分野で研究成果を出しており、HIN に関する研究においても研究成果 [6,7,8,12,13] を上げている。

## 【主要な論文リスト】

- [1] [Takahiro Komamizu](#), Toshiyuki Amagasa, Hiroyuki Kitagawa, "Facet-value Extraction Scheme from Textual Contents in XML Data", IJWIS, Vol. 11, Iss. 3, pp.270-290, 2015
- [2] [Takahiro Komamizu](#), Toshiyuki Amagasa, Hiroyuki Kitagawa, "Faceted Navigation Framework for XML Data", IJWIS, Vol. 8, Iss. 4, pp.348-370, 2012
- [3] 駒水 孝裕, 天笠 俊之, 北川 博之, 異種 XML データに対するファセット検索手法の提案, 情報処理学会研究報告 (DD), 2009-DD-073(7), pp. 1-8, 2009 年 9 月
- [4] [Takahiro Komamizu](#), Yuto Yamaguchi, Toshiyuki Amagasa, Hiroyuki Kitagawa, "FACTUS: Faceted Twitter User Search Using Twitter Lists", in Proc. WISE 2011, pp.343-344, 2011
- [5] [Takahiro Komamizu](#), Mariko Kamie, Kazuhiro Fukui, Toshiyuki Amagasa, Hiroyuki Kitagawa, "A Scheme of Fragment-Based Faceted Image Search", in Proc. DEXA 2012, pp.450-457, 2012
- [6] [Takahiro Komamizu](#), Toshiyuki Amagasa, Hiroyuki Kitagawa, "SPOOL: A SPARQL-based ETL Framework for OLAP over Linked Data", in Proc. iiWAS 2015, pp.49:1-10, 2015
- [7] [Takahiro Komamizu](#), "Learning Interpretable Entity Representation in Linked Data", in Proc. DEXA 2018, pp. 153-168, 2018
- [8] Takahiro Komamizu, "Graph Analytical Re-ranking for Entity Search", in Proc. EYRE@CIKM 2018, (採択済み)
- [9] [Takahiro Komamizu](#), Jin Nakazawa, Toshiyuki Amagasa, Hiroyuki Kitagawa, Hideyuki Tokuda, "Analytical Toolbox for Smart City Applications: Garbage Collection Log Use Case", in Proc. IWSC@IEEE Big Data 2017, pp.4105-4110, 2017
- [10] [Takahiro Komamizu](#), Toshiyuki Amagasa, Salman Ahmed Shaikh, Hiroaki Shiokawa, Hiroyuki Kitagawa, "SOLA: Stream OLAP-based Analytical Framework for Roadway Maintenance", in Proc. MEDES 2017, pp.35-41, 2017
- [11] [Takahiro Komamizu](#), Toshiyuki Amagasa, Hiroyuki Kitagawa, "Visual Spatial-OLAP for Vehicle Recorder Data on Micro-sized Electric Vehicles", in Proc. IDEAS 2016, pp.358-363, 2016
- [12] Hiroyoshi Ito, [Takahiro Komamizu](#), Toshiyuki Amagasa, Hiroyuki Kitagawa, "Community Detection and Correlated Attribute Cluster Analysis on Multi-Attributed Graphs", in Proc. DARLI-AP@EDBT 2018, pp.2-9, 2018
- [13] Hiroyoshi Ito, [Takahiro Komamizu](#), Toshiyuki Amagasa, Hiroyuki Kitagawa, "Network-Word Embedding for Dynamic Text Attributed Networks", in Proc. SCSN@ICSC 2018, pp.334-339, 2018

## 背景の補足：異種情報ネットワークと Meta-path

➤ **異種情報ネットワーク (Heterogeneous Information Network, HIN)** は、複数種類のオブジェクトから構成されるネットワークである。図 1 に HIN の例を示す。図 1 では、人、スポーツ、場所、食事の 4 種類のオブジェクトが含まれている。オブジェクト間に張られる辺はそれらの間に関係があることを示す。例えば、人「A」と食事「ひつまぶし」には関係がある。この例では、「好物」の関係を表現している。

➤ **Meta-path** は HIN 上に定義されるオブジェクトの種類間の関係の並びである。例えば、図 1 の例において、Meta-path 「人→スポーツ→人」は同じスポーツをする人の関係を表現できる。Meta-path 「人→食事→場所→食事→人」は同じ場所で名物になっている食事が好物である人の関係を表現できる。このような関係の記述から Meta-path はオブジェクト間の関連性の評価に利用される [1,2,3]。オブジェクト間の関連性を評価することで、食事の推薦や友人の推薦など様々なアプリケーションへの応用が可能となる。

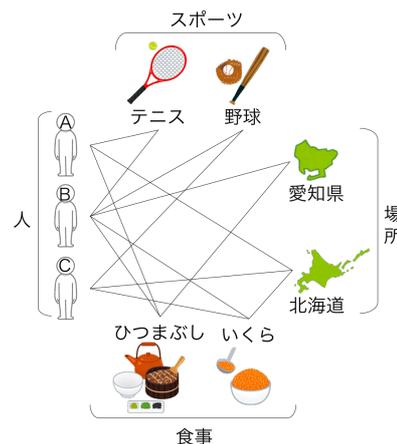


図 1：HIN の例

## 研究の内容の補足：研究のアプローチ手順

オブジェクト同士の関連度から Meta-path を逆算する際に、どのような教師データがあれば実現できるかを明らかにすることが重要な課題である。本研究では、教師データの質および量と Meta-path を逆算する精度との関係を明らかにする。

- **教師データの質**：Meta-path を基に計算された関連度と同値を与えた場合に、あらゆる Meta-path を試行することで正確な Meta-path が逆算できることが期待される。しかしながら、Meta-path を基に計算された関連度をそのまま得られるという前提は現実的ではない。故に、関連度の値にノイズを乗せた場合や、同関連度に基づく関連性ランキングが与えられた場合を考えるほうが現実的である。本研究では、ノイズの大きさやランキングの入れ替えなどをコントロールし、教師データの質と Meta-path の逆算精度の関係を明らかにする。
- **教師データの量**：一般に、予測モデルを構築する際の教師データの量は多ければ多いほどよいとされている。Meta-path の逆算においても、すべてのオブジェクト対について Meta-path を用いた関連度の正確な値を教師データとして用意すれば、Meta-path の逆算は確実にできる。しかしながら、十分な数の関連度を、使用された Meta-path の知識なしに用意することは困難である。故に、一定数間引いた教師データを仮定することが現実的である。本研究では、教師データの数と Meta-path の逆算精度の関係を明らかにする。
- **質と量の関係**：本研究では、教師データの質および量について、独立に Meta-path 逆算精度との関係性を明らかにするとともに、質と量を同時に変えたときの Meta-path 逆算の精度との関係性を明らかにする。これを実現することにより、人間が関連すると考えるオブジェクト同士の組合せやその関連度から、その関連度を説明するような Meta-path を導出することが可能となる。更に、導出した Meta-path を用いて他のオブジェクト同士の関連性を測ることができるようになる。

## 【参考文献】

- [1] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu, “PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks,” PVLDB, vol. 4, no. 11, pp. 992–1003, 2011.
- [2] Y. Dong, N. V. Chawla, and A. Swami, “metapath2vec: Scalable Representation Learning for Heterogeneous Networks,” in SIGKDD 2017, 2017, pp. 135–144.
- [3] T. Fu, W. Lee, and Z. Lei, “HIN2Vec: Explore Meta-paths in Heterogeneous Information Networks for Representation Learning,” in CIKM 2017, 2017, pp. 1797–1806.