

令和 2 年 6 月 11 日現在

機関番号：14301

研究種目：基盤研究(A)（一般）

研究期間：2015～2019

課題番号：15H01704

研究課題名（和文）ビッグデータ時代の複雑構造データを扱う機械学習法の研究

研究課題名（英文）Machine Learning for Complex-Structured Data

研究代表者

鹿島 久嗣（Kashima, Hisashi）

京都大学・情報学研究科・教授

研究者番号：80545583

交付決定額（研究期間全体）：（直接経費） 33,100,000円

研究成果の概要（和文）：本研究課題では、グラフ構造などの複雑な構造をもつデータや、分析過程において人間によって生成されたデータを効果的に扱う手法について研究を行った。特に、近年著しく発展しているグラフ構造を対象とした深層学習法の開発を行った。一方、機械学習だけでは解決が困難な、高い専門性や抽象的な判断が必要な課題に対して、人間が判断したデータを適切に処理することで、正しい判断を導く数々の方法論を開発した。また、以上で開発した手法を、化学や生物学・材料科学などの科学分野を中心としたその応用に取り組み、その有効性を示した。

研究成果の学術的意義や社会的意義

本研究は、近年学術界・産業界において期待されている深層学習などの機械学習・データ解析技術の適用可能性を大きく広げるとともに、人工知能だけでは解決が困難な問題に対して人間の知能や判断を組み合わせること、より適切な意思決定を行うことを可能とするための基盤技術を開発するものである。

研究成果の概要（英文）：We studied effective data analysis methods for complex-structured data such as graph-structured data and human-generated data. In particular, we developed new models and techniques for deep learning for graphs. On the other hand, we developed various methods to integrate human judges to make better decisions on difficult problems that machine intelligence alone cannot solve. We also applied the developed methods to various scientific applications in chemistry, biology, and material science, and showed their promise.

研究分野：人工知能

キーワード：人工知能 機械学習 データサイエンス ヒューマンコンピューテーション クラウドソーシング

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。

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

近年の世界的なビッグデータの潮流は産業界だけでなく学术界にも大きな影響を与えており、科学研究においてもデータを主軸とした考え方へのパラダイムシフトが起こっている。ビッグデータの扱いにおいて中心となる技術は、機械学習をはじめとする自動的なデータ解析技術である。近年の機械学習技術の進展は目覚ましく、先端的な Web 企業が中核技術として機械学習技術を活用している。一方で、これまで人間のほうが有利と思われてきた知的で複雑な実世界タスクにおいても、機械がもっとも熟練した人間を凌駕しうることが示されている。やはり、これらのシステムにおいても機械学習が中核技術として用いられていることはよく知られた事実である。しかしながら、産業界を中心にデータサイエンティストと呼ばれるデータ解析従事者の不足が叫ばれていることからわかるように、データ解析がその属人性・労働集約性からいまだ解放されていないこともまた事実であり、一層の研究が必要である。ビッグデータの特徴づける特徴として、しばしば量・速度・多様性が挙げられ、技術的には既存の方法論の大規模化や高速化に注力されることが多いが、一方で、多様で複雑な形式のデータをいかに扱うか、あるいはデータ量をいかに確保するかといった観点からの議論は十分でなかった。

2. 研究の目的

本研究は、近年学术界・産業界において期待されている深層学習などの機械学習・データ解析技術の適用可能性を大きく広げるとともに、人工知能だけでは解決が困難な問題に対して人間の知能や判断を組み合わせることで、より適切な意思決定を行うことを可能とするための基盤技術を開発することを目指すものである。

ビッグデータ研究においてしばしば看過されがちであるのが、多様で複雑なデータ形式への対処と、人手の作業を必要とする大量メタデータの確保である。本研究はこのような観点から、グラフなどの複雑な構造をもつビッグデータを対象とした機械学習法の確立を目指す。また、計算機と人間との協働プラットフォームとして普及しつつあるクラウドソーシングを利用したメタデータ収集と、これに基づく機械学習との融合を実現する。さらに上記の技術の応用として、膨大な有機低分子化合物の候補から新規薬剤など有用化合物候補を発見する開発プロセスの効率化に取り組む。

3. 研究の方法

複雑な構造データからの効率的で効果的な機械学習法、クラウドソーシングによって収集されたメタデータからの機械学習法、また、様々な実世界応用を行う。まず、複雑な構造データからの機械学習法においては、グラフや、さらに高い表現力をもつテンソルやハイパーグラフなどデータ形式を対象として、表現力が高く効率的・高精度な解析手法を確立する。とりわけ、近年著しい発展を遂げている深層学習を基礎に、強力な構造データの学習手法を開発する。

クラウドソーシングによって収集されたメタデータからの機械学習においては、これを効果的に行う技術を確立する。特に、従来は単純な択一選択タスクを対象としてきたのに対し、より複雑な、ランキングや非定型回答タスクなどを扱う手法を開発する。

応用としては、医療や材料・有機化学等の分野において、先進的な機械学習技術の応用や、機械学習と人間が融合した系による意思決定支援など、領域専門家や企業との協力のもと現実の課題・データを用いたケーススタディ等を行う。

4. 研究成果

本研究課題では、グラフ構造などの複雑な構造をもつデータや、分析過程において人間によって生成されたデータを効果的に扱う手法について研究を行った。特に、近年著しく発展しているグラフ構造を対象とした深層学習法の開発を行った。一方、機械学習だけでは解決が困難な、高い専門性や抽象的な判断が必要な課題に対して、人間が判断したデータを適切に処理することで、正しい判断を導く数々の方法論を開発した。また、以上で開発した手法を、化学や生物学・材料科学などの科学分野を中心としたその応用に取り組み、その有効性を示した。以下具体的な研究成果について述べる。

まず、構造をもつデータを対象とした基礎的な機械学習手法として、グラフの構造予測を、その接続行列に着目した行列分解によって行う新しい手法構造予測法を開発し、この手法が高い効果を発揮する条件として、グラフがある種の疎性をもつことを発見した。一方、近年著しく発展しているグラフ深層学習法において、モデルが行う判断の根拠が不明であることを解決するため、化合物等のグラフ構造のどの部分構造が判断の根拠になっているかを提示する頑強な手法を開発し、その有効性を示した。また、グラフ深層学習の理論的考察として、グラフ深層学習モデルの近似能力の解析と、より能力の高いモデルのクラスを提案した。さらに、グラフよりも一般的な階層構造をもった「グラフのグラフ」に対する深層学習法を開発し、これを化合物間相互作用の予測に応用する研究を行った。

人間によって与えられたメタデータの品質を高める方法についての研究としては、階層的な分類体系をもつようなメタデータに対して、項目反応理論の段階反応モデルをもとに成果品質のばらつきを吸収し、より精度の高いメタデータを得るための方法を開発した。また、2つの対象同志を比較することによって得られるようなメタデータを効率的に収集して、より正確な順位を推定するためのデータ収集法と統計的予測法を開発した。さらに、第一段階目で成果物を生

成し、第二段階目で成果物の評価を行うような多段のプロセスにおいて、作成者の能力や評価者のバイアス明示的に考慮しながら、成果物品質を推定する手法を、Web ページのランキング手法である HITS を拡張することで開発した。その他にも、効率的なメタデータ収集を実現するために、適切な人に適切なタスクを推薦するための推薦手法を開発した。

一方、一層困難な課題として、専門知識を要する困難な問題に対する新しい統計的意見統合手法として、複数の問題の組み合わせを「超問題」として考え、この上で意見統合を行うことで、専門家の意見を強調できるようになる新しい手法を開発した。さらに、抽象的な問題に対する解決法やデザインなどの入出力がはっきりしない抽象的な課題においては、機械による直接的な扱いが困難であるが、これに対して、多数の人間によって対象の優先度付けや類似性評価が実施されたデータをもとに、これを統合的に用いることで、優先度と類似度をともに反映した、効率的な低次元空間表現の獲得手法を開発した。より機械と人間が密接に協力する系として、判断に有効な特徴を人間が作成し、これに基づき機械学習モデリングを行うような人間参加型の機械学習系において、低コストで有用な特徴抽出を行う手法を、機械学習のブースティングの枠組みを用いて実現した。

応用においては、行政機関等からオープンデータとして公開されるデータの再利用性を高めるために、画像として公開されているそのままでは再利用性の低いデータから、クラウドソーシングを用いて再利用可能なデータに変換する試みを行い、タスク依頼の方法や集計方法を工夫することで、高い品質で変換が行えることを確認した。また、高度な専門性を必要とする課題である、化合物の合成可能性判定問題において、準専門家の判定を統計的に判定することによって専門家に匹敵する精度で判断を行えることを示した。

その他にも、センサーネットワークにおいて通信量を削減し効率的に分散学習を行う仕組みや、深層学習を用いた短期的な気象予測、テンソル分解による関係分析手法を用いた材料科学における推薦システムなど、様々な先進的応用を行った。

5. 主な発表論文等

〔雑誌論文〕 計12件（うち査読付論文 12件／うち国際共著 0件／うちオープンアクセス 7件）

1. 著者名 Atsuto Seko, Hiroyuki Hayashi, Hisashi Kashima, Isao Tanaka	4. 巻 2(1)
2. 論文標題 Matrix- and Tensor-based Recommender Systems for the Discovery of Currently Unknown Inorganic Compounds	5. 発行年 2018年
3. 雑誌名 Physical Review Materials	6. 最初と最後の頁 013805.1-8
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.1103/PhysRevMaterials.2.013805	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Jiyi Li, 馬場 雪乃, 鹿島 久嗣	4. 巻 17-J
2. 論文標題 超問題：専門知識を要するクラウドソーシングタスクの回答統合	5. 発行年 2019年
3. 雑誌名 日本データベース学会和文論文誌	6. 最初と最後の頁 1-8
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Yukino Baba, Tetsu Isomura, Hisashi Kashima	4. 巻 80
2. 論文標題 Wisdom of Crowds for Synthetic Accessibility Evaluation	5. 発行年 2018年
3. 雑誌名 Journal of Molecular Graphics and Modelling	6. 最初と最後の頁 217-223
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.1016/j.jmgs.2018.01.011	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Sho Yokoi, Hiroshi Kajino, Hisashi Kashima	4. 巻 25
2. 論文標題 Link Prediction in Sparse Networks by Incidence Matrix Factorization	5. 発行年 2017年
3. 雑誌名 Journal of Information Processing	6. 最初と最後の頁 477-485
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.2197/ipsjip.25.477	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Satoshi Oyama, Yukino Baba, Ikki Ohmukai, Hiroaki Dokoshi, Hisashi Kashima	4. 巻 2
2. 論文標題 Crowdsourcing Chart Digitizer: Task Design and Quality Control for Making Legacy Open Data Machine-Readable	5. 発行年 2016年
3. 雑誌名 International Journal of Data Science and Analytics	6. 最初と最後の頁 45-60
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Yukino Baba, Kei Kinoshita, Hisashi Kashima	4. 巻 58
2. 論文標題 Participation Recommendation System for Crowdsourcing Contests	5. 発行年 2016年
3. 雑誌名 Expert Systems with Applications	6. 最初と最後の頁 174-183
掲載論文のDOI (デジタルオブジェクト識別子) https://doi.org/10.1016/j.eswa.2016.04.010	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Naoki Otani, Yukino Baba, Hisashi Kashima	4. 巻 58
2. 論文標題 Quality Control of Crowdsourced Classification Using Hierarchical Class Structures	5. 発行年 2016年
3. 雑誌名 Expert Systems with Applications	6. 最初と最後の頁 155-163
掲載論文のDOI (デジタルオブジェクト識別子) https://doi.org/10.1016/j.eswa.2016.04.009	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 梶村 俊介, 馬場 雪乃, 梶野 洸, 鹿島 久嗣	4. 巻 31
2. 論文標題 列挙型クラウドソーシングタスクのための品質管理法	5. 発行年 2016年
3. 雑誌名 人工知能学会論文誌	6. 最初と最後の頁 1-9
掲載論文のDOI (デジタルオブジェクト識別子) http://doi.org/10.1527/tjsai.K-F79	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 則 のぞみ, ボレガラ ダヌシカ, 鹿島 久嗣	4. 巻 30
2. 論文標題 接続行列埋め込みに基づく複数種類の多項関係の同時予測	5. 発行年 2015年
3. 雑誌名 人工知能学会論文誌	6. 最初と最後の頁 459-465
掲載論文のDOI (デジタルオブジェクト識別子) http://doi.org/10.1527/tjsai.30.459	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

1. 著者名 Naoki Otani, Yukino Baba, Hisashi Kashima	4. 巻 1
2. 論文標題 Quality Control for Crowdsourced Hierarchical Classification	5. 発行年 2015年
3. 雑誌名 Proceedings of the 2015 IEEE International Conference on Data Mining (ICDM)	6. 最初と最後の頁 937-942
掲載論文のDOI (デジタルオブジェクト識別子) http://doi.org/10.1016/j.eswa.2016.04.009	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Satoshi Oyama, Yukino Baba, Ikki Ohmukai, Hiroaki Dokoshi, Hisashi Kashima	4. 巻 1
2. 論文標題 From One Star to Three Stars: Upgrading Legacy Open Data Using Crowdsourcing	5. 発行年 2015年
3. 雑誌名 Proceedings of the 2015 International Conference on Data Science and Advanced Analytics (DSAA)	6. 最初と最後の頁 1-9
掲載論文のDOI (デジタルオブジェクト識別子) http://doi.org/10.1109/DSAA.2015.7344801	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Shunsuke Kajimura, Yukino Baba, Hiroshi Kajino, Hisashi Kashima	4. 巻 1
2. 論文標題 Quality Control for Crowdsourced POI Collection	5. 発行年 2015年
3. 雑誌名 Proceedings of the 19th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)	6. 最初と最後の頁 255-267
掲載論文のDOI (デジタルオブジェクト識別子) http://10.1007/978-3-319-18032-8_20	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計18件（うち招待講演 0件 / うち国際学会 13件）

1. 発表者名 Hirotaka Akita, Kosuke Nakago, Tomoki Komatsu, Yohei Sugawara, Shin-ichi Maeda, Yukino Baba, Hisashi Kashima
2. 発表標題 BayesGrad: Explaining Predictions of Graph Convolutional Networks
3. 学会等名 Proceedings of the 25th International Conference on Neural Information Processing (ICONIP) (国際学会)
4. 発表年 2018年

1. 発表者名 Ryoma Sato, Takehiro Yamamoto, Hisashi Kashima
2. 発表標題 Short-term Precipitation Prediction with Skip-connected PredNet
3. 学会等名 Proceedings of the 27th International Conference on Artificial Neural Networks (ICANN) (国際学会)
4. 発表年 2018年

1. 発表者名 Jiyi Li, Hisashi Kashima
2. 発表標題 Incorporating Worker Similarity for Label Aggregation in Crowdsourcing
3. 学会等名 Proceedings of the 27th International Conference on Artificial Neural Networks (ICANN) (国際学会)
4. 発表年 2018年

1. 発表者名 Jiyi Li, Yukino Baba, Hisashi Kashima
2. 発表標題 Simultaneous Clustering and Ranking from Pairwise Comparisons
3. 学会等名 Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI) (国際学会)
4. 発表年 2018年

1 . 発表者名 Guoxi Zhang, Tomoharu Iwata, Hisashi Kashima
2 . 発表標題 On Reducing Dimensionality of Labeled Data Efficiently
3 . 学会等名 Proceedings of the 22nd Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) (国際学会)
4 . 発表年 2018年

1 . 発表者名 Jiyi Li, Yukino Baba, Hisashi Kashima
2 . 発表標題 Hyper Questions: Unsupervised Targeting of a Few Experts in Crowdsourcing
3 . 学会等名 Proceeding of the 26th ACM International Conference on Information and Knowledge Management (CIKM) (国際学会)
4 . 発表年 2017年

1 . 発表者名 Kosuke Yoshimura, Yukino Baba and Hisashi Kashima
2 . 発表標題 Quality Control for Crowdsourced Multi-Label Classification using RAKEL
3 . 学会等名 Proceeding of the 24th International Conference on Neural Information Processing (ICONIP) (国際学会)
4 . 発表年 2017年

1 . 発表者名 Jiyi Li, Tomohiro Arai, Yukino Baba, Hisashi Kashima, Shotaro Miwa
2 . 発表標題 Distributed Multi-task Learning for Sensor Network
3 . 学会等名 roceeding of the European Conference on Machine Learning & Principles and Practice of Knowledge Discovery in Databases (ECML PKDD) (国際学会)
4 . 発表年 2017年

1 . 発表者名 Guoxi Zhang, Tomoharu Iwata, Hisashi Kashima
2 . 発表標題 Robust Multi-view Topic Modeling by Incorporating Detecting Anomalies
3 . 学会等名 Proceeding of the European Conference on Machine Learning & Principles and Practice of Knowledge Discovery in Databases (ECML PKDD) (国際学会)
4 . 発表年 2017年

1 . 発表者名 Jiuding Duan, Jiyi Li, Yukino Baba, Hisashi Kashima
2 . 発表標題 A Generalized Model for Multidimensional Intransitivity
3 . 学会等名 Proceedings of the 21st Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) (国際学会)
4 . 発表年 2017年

1 . 発表者名 Yukino Baba, Tomoumi Takase, Kyohei Atarashi, Satoshi Oyama, Hisashi Kashima
2 . 発表標題 Data Analysis Competition Platform for Educational Purposes: Lessons Learned and Future Challenges
3 . 学会等名 Proceedings of the 8th Symposium on Educational Advances in Artificial Intelligence (EAAI) (国際学会)
4 . 発表年 2018年

1 . 発表者名 Takeru Sunahase, Yukino Baba, Hisashi Kashima
2 . 発表標題 Pairwise HITS: Quality Estimation from Pairwise Comparisons in Creator-Evaluator Crowdsourcing Process
3 . 学会等名 Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI) (国際学会)
4 . 発表年 2017年

1. 発表者名 Sho Yokoi, Hiroshi Kajino, Hisashi Kashima
2. 発表標題 Link Prediction by Incidence Matrix Factorisation
3. 学会等名 Proceedings of the 22nd European Conference on Artificial Intelligence (ECAI) (国際学会)
4. 発表年 2016年

1. 発表者名 高濱 隆輔, 神鷲 敏弘, 鹿島 久嗣
2. 発表標題 漸進比較法によるランキング推定
3. 学会等名 人工知能学会全国大会
4. 発表年 2015年

1. 発表者名 横井 祥, 梶野 洸, 鹿島 久嗣
2. 発表標題 接続行列分解による関係予測
3. 学会等名 人工知能学会全国大会
4. 発表年 2015年

1. 発表者名 大谷 直樹, 馬場 雪乃, 鹿島 久嗣
2. 発表標題 クラウドソーシングにおける階層的分類タスクの品質管理手法
3. 学会等名 人工知能学会全国大会
4. 発表年 2015年

1. 発表者名 馬場 雪乃, 木下 慶, 鹿島 久嗣
2. 発表標題 コンテスト型クラウドソーシングにおける勝敗予測.
3. 学会等名 人工知能学会全国大会
4. 発表年 2015年

1. 発表者名 梶野 洸, 荒井 ひろみ, 佐久間 淳, 鹿島 久嗣
2. 発表標題 クラウドソーシングにおけるプライバシー保護タスク割り当て.
3. 学会等名 人工知能学会全国大会
4. 発表年 2015年

〔図書〕 計1件

1. 著者名 鹿島久嗣, 小山 聡, 馬場雪乃	4. 発行年 2016年
2. 出版社 講談社	5. 総ページ数 128
3. 書名 ヒューマンコンピューテーションとクラウドソーシング	

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考