様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月 ２３ 日現在

機関番号：１４３０１
研究種目：基盤研究(B)（一般）
研究期間：2019～2021
課題番号：１９Ｈ０４１６９

研究課題名（和文）複数のテンソルからの効率的なデータ構造推定

研究課題名（英文）Efficient estimation of data structure from multiple tensors

研究代表者
馬見塚　拓（Mamitsuka, Hiroshi）

京都大学・化学研究所・教授

研究者番号：００３４６１０７

交付決定額（研究期間全体）：（直接経費）　　13,200,000円

研究成果の概要（和文）：現代のビッグデータには、以下の２つの特徴がある。１，高次元データ（例えば、ユーザｘ商品ｘ時間という購買データ）、２、モードを共有する複数データ（例えば、上記高次元データに、ユーザ間の情報を加えれば、ユーザというモードを共有する）。そこで、モードを共有する複数高次元データから、内在する因子を効率的に抽出するための、なるたけ一般的な手法を構築した。具体的には、複数高次元データを表現可能なノルムを、効率的な学習が可能なように開発し、その性質を解析するとともに、性能の優位性を実験的に示した。また、開発経験を活かし、高次元データと行列に関する様々な問題への解決手法を提案し、性能を応用において実証した。

研究成果の学術的意義や社会的意義
現在、科学、工学、ビジネス等、社会の様々な場所で多様なデータが生まれるビッグデータの時代を迎えており、データに内在する構造を効率的に理解する技術の整備が喫緊の課題である。特に、現代では、モードを共有する複数高次元データがよく見られる。このようなデータに対し、本研究は、データに内在する構造を理解する一般的な枠組みと実際の事例を与える。また、本研究で得られた知見を使うことにより、関係するデータに対しても、効率的で精度の高い手法の構築が可能なことを示した。以上から、本研究は、現代に増大しつつある、より複雑な複数の関係データに対するデータ科学および機械学習手法開発の進展に大きく貢献する。

研究成果の概要（英文）：Modern big data has the following two properties: 1) High-dimensional data (i.e. tensors, for example, purchase data, like users x items x time), 2) multiple tensors sharing the same mode (for example, if we have users' mutual information with the above tensor, these two information sources share the mode of users). Thus, we developed a general method, which can extract latent components hidden in given multiple tensors, which share one or more modes. In reality, we design a norm to represent given multiple tensors, so that this norm allows to formulate optimization (learning) problems to be solved efficiently. At the same time, we analyzed the properties of the norm theoretically, and also evaluated the empirical performence of the norms in the settings of both synthetic and real data. Also, by using the experience we had cultivated through the development of the norms, we developed numerous solutions for the problems related with high-dimensional data, i.e. tensors, and matrices.

研究分野：機械学習

キーワード：機械学習

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

現在、科学、工学、ビジネス等、社会の様々な場所で多様なデータが生まれるビッグデータの時代を迎えており、データに内在する構造を効率的に理解する「データ科学」技術の整備が喫緊の課題である。ビッグデータが持つ多くの側面の中で、本研究は以下の 2 つの重要な側面に着目する：

1) 高次元データ：従来、最も一般的なデータは行列である。例えば、購買データは、行がユーザ、列が商品の行列である。これに、例えば、時間を次元に加えると、「ユーザ X 商品 X 時間」の 3 次元のテンソルとなる。現在、テンソルのような高次元データが増加している。
2) モードを共有する複数データ：従来、入力データ（行列 X）は 1 つで、上述の購買データは典型例である。現在さらに、例えばソーシャル・ネットワーキング・サーヴィス(SNS)でユーザ間リンク(隣接行列 W)が得られる。X と W は、ユーザのモードを共有する。つまり、X の内在構造を X のみならず、与えられた複数行列全てから包括的に理解できる。X に対し W は補助情報と呼ばれる。

これら、2 つのデータ各々に対しては、様々な研究がなされ、解決手法がある程度確立している。以下、内在する因子を取り出すことができる分解（あるいは教師なし学習）の観点から例をしめす。

1) 高次元データ：長い歴史を持つテンソル分解により構造推定でき、特に 2 つの分解手法が代表的である。
   - CANDECOMP/PARAFAC (CP)分解
   - Tucker 分解
2) モードを共有する複数行列 : 典型的な解法は、X を 2 つの部分行列 U と V に行列分　解する際に、この行列分解だけを目的関数するのではなく、モードを共有する U または V と W を使った正則化項を含めた全体を目的関数とする。この解決方法は「協調行列分解 (collaborative matrix factorization)」と総称されている。

しかし、上記、2 つのデータを同時に考慮した状況、すなわち「モードを共有する複数高次元データ」に関しては、内在する因子を取り出すために提案されたこれまでの方法には、以下の問題点があり、一般的な方法が確立していない。

1. 非効率：定式化された最適化問題は凸問題ではない等、最適解を得るための効率的な解法ではない。
2. 適用データの限定：任意の複数高次元データに適用ができず、制限がある。
3. 簡単な拡張 : モードを共有する複数行列の分解からの比較的単純な拡張が多く、高次元データすなわちテンソルの性質を考慮して設計されていない。

２．研究の目的

以上の背景から、本研究の目的は以下の 2 つである。

手法構築 : モードを共有する複数高次元データの因子構造を効率的に理解する、より一般的手法の構築。

理論解析：上記手法を理論的に支持するための学習誤差やテスト誤差の理論的な解析・解明。

上記 2 つの目的の実現を目指す本研究は以下の特徴を持つ。

1. 効率性：高次元データは行列に比べ要素数が多く、加えて、複数テンソルにより、さらに多くなる。従って、効率性はより重要である。効率的な解法が存在し得る最適化問題として定式化する。

2. データ自由度：入力の高次元データの数や次元になるたけ制限のない自由度の高い手法を構築する。

3. 実高次元データの性質を考慮：高次元データは欠損値の割合が非常に高くなる。すなわちスパース性が高くなる。テンソルの実際的な性質を考慮した手法を構築する。

３．研究の方法

協調行列分解で正則化項が使われたように、複数高次元データにおいてもノルムの構築が重要である。特に、複数の項が目的関数にあるよりも、少ない数の項で構成できれば、より計算が効率的になる。そこで、本研究では、複数高次元データを 1 つのノルムで表現し、ノルム内のパラメータをデータから推定する。また、複数高次元データに対するノルムの構築で得られた知見を活かし、特に生命科学を中心とした高次元データの解析を行い、それらの成果を論文にまとめていく。

４．研究成果

複数高次元データから因子を効率的に得るためのスケーラブルなノルムの設計を行い、理論・実験両面からその有効性を示した（*Neural Computation*, 2020)。また、単一の高次元データを構成する際に、計算効率性の高いノルムの設計を行い、理論・実験両面からその有効性を示した（*Machine Learning*, 2021)。また、行列分解に基づく、スパース性を考慮したマルチビューマルチタスク学習手法を複数構築した（*IJCAI*, 2019 で 2 報)。同時に、行列分化において、補助情報の隣接行列がスパースな場合に、学習精度を向上させる手法の構築を行った(*AAAI*, 2020)。その他、生命科学での複数のモードを共有する複数行列の協調行列分解を行い、個別化医療の推進に貢献した（*Briefings in Bioinformatics*, 2021; *IEEE TCBB*, 2023)。

５．主な発表論文等

〔雑誌論文〕　計27件（うち査読付論文　22件／うち国際共著　13件／うちオープンアクセス　21件）

１．著者名 Nguyen Dai Hai、Nguyen Canh Hao、Mamitsuka Hiroshi	４．巻 35
２．論文標題 ADAPTIVE: leArning DAta-dePendenT, concIse molecular VEctors for fast, accurate metabolite identification from tandem mass spectra	５．発行年 2019年
３．雑誌名 Bioinformatics	６．最初と最後の頁 i164～i172
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bioinformatics/btz319	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 You Ronghui、Yao Shuwei、Xiong Yi、Huang Xiaodi、Sun Fengzhu、Mamitsuka Hiroshi、Zhu Shanfeng	４．巻 47
２．論文標題 NetGO: improving large-scale protein function prediction with massive network information	５．発行年 2019年
３．雑誌名 Nucleic Acids Research	６．最初と最後の頁 W379～W387
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/nar/gkz388	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 該当する

１．著者名 Sun Lu、Nguyen Canh Hao、Mamitsuka Hiroshi	４．巻 -
２．論文標題 Fast and Robust Multi-View Multi-Task Learning via Group Sparsity	５．発行年 2019年
３．雑誌名 Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI 2019)	６．最初と最後の頁 3499-3505
掲載論文のDOI（デジタルオブジェクト識別子） 10.24963/ijcai.2019/485	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 Sun Lu、Nguyen Canh Hao、Mamitsuka Hiroshi	４．巻 -
２．論文標題 Multiplicative Sparse Feature Decomposition for Efficient Multi-View Multi-Task Learning	５．発行年 2019年
３．雑誌名 Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI 2019)	６．最初と最後の頁 3506-3512
掲載論文のDOI（デジタルオブジェクト識別子） 10.24963/ijcai.2019/486	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 You, R., Dai, S., Zhang, Z., Mamitsuka, H. and Zhu, S.	4．巻 -
2．論文標題 AttentionXML: Label Tree-based Attention-Aware Deep Model for High-Performance Extreme Multi-Label Text Classification.	5．発行年 2019年
3．雑誌名 Proceedings of the 33rd Annual Conference on Neural Information Processing Systems (NeurIPS 2019)	6．最初と最後の頁 5820-5830
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 該当する

1．著者名 Dai Suyang、You Ronghui、Lu Zhiyong、Huang Xiaodi、Mamitsuka Hiroshi、Zhu Shanfeng	4．巻 36
2．論文標題 FullMeSH: improving large-scale MeSH indexing with full text	5．発行年 2019年
3．雑誌名 Bioinformatics	6．最初と最後の頁 1533～1541
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bioinformatics/btz756	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 該当する

1．著者名 Wimalawarne Kishan、Yamada Makoto、Mamitsuka Hiroshi	4．巻 32
2．論文標題 Scaled Coupled Norms and Coupled Higher-Order Tensor Completion	5．発行年 2020年
3．雑誌名 Neural Computation	6．最初と最後の頁 447～484
掲載論文のDOI（デジタルオブジェクト識別子） 10.1162/neco_a_01254	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 Nakamura Atsuyoshi、Takigawa Ichigaku、Mamitsuka Hiroshi	4．巻 34
2．論文標題 Efficiently Enumerating Substrings with Statistically Significant Frequencies of Locally Optimal Occurrences in Gigantic String	5．発行年 2020年
3．雑誌名 Proceedings of the AAAI Conference on Artificial Intelligence	6．最初と最後の頁 5240～5247
掲載論文のDOI（デジタルオブジェクト識別子） 10.1609/aaai.v34i04.5969	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 Strahl Jonathan、Peltonen Jaakko、Mamitsuka Hirsohi、Kaski Samuel	4．巻 34
2．論文標題 Scalable Probabilistic Matrix Factorization with Graph-Based Priors	5．発行年 2020年
3．雑誌名 Proceedings of the AAAI Conference on Artificial Intelligence	6．最初と最後の頁 5851～5858
掲載論文のDOI（デジタルオブジェクト識別子） 10.1609/aaai.v34i04.6043	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 該当する

1．著者名 Liu Lizhi、Huang Xiaodi、Mamitsuka Hiroshi、Zhu Shanfeng	4．巻 36
2．論文標題 HPOLabeler: improving prediction of human protein?phenotype associations by learning to rank	5．発行年 2020年
3．雑誌名 Bioinformatics	6．最初と最後の頁 4180～4188
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bioinformatics/btaa284	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 該当する

1．著者名 Nguyen Hao Canh、Mamitsuka Hiroshi	4．巻 43
2．論文標題 Learning on Hypergraphs with Sparsity	5．発行年 2020年
3．雑誌名 IEEE Transactions on Pattern Analysis and Machine Intelligence	6．最初と最後の頁 2710-2722
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TPAMI.2020.2974746	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 You Ronghui、Liu Yuxuan、Mamitsuka Hiroshi、Zhu Shanfeng	4．巻 37
2．論文標題 BERTMeSH: deep contextual representation learning for large-scale high-performance MeSH indexing with full text	5．発行年 2020年
3．雑誌名 Bioinformatics	6．最初と最後の頁 684～692
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bioinformatics/btaa837	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 該当する

1．著者名 Wimalawarne Kishan、Mamitsuka Hiroshi	4．巻 110
2．論文標題 Reshaped tensor nuclear norms for higher order tensor completion	5．発行年 2021年
3．雑誌名 Machine Learning	6．最初と最後の頁 507～531
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s10994-020-05927-y	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 Nguyen Duc Anh、Nguyen Canh Hao、Mamitsuka Hiroshi	4．巻 22
2．論文標題 A survey on adverse drug reaction studies: data, tasks and machine learning methods	5．発行年 2019年
3．雑誌名 Briefings in Bioinformatics	6．最初と最後の頁 164～177
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bib/bbz140	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 Guvenc Paltun Betul、Mamitsuka Hiroshi、Kaski Samuel	4．巻 22
2．論文標題 Improving drug response prediction by integrating multiple data sources: matrix factorization, kernel and network-based approaches	5．発行年 2019年
3．雑誌名 Briefings in Bioinformatics	6．最初と最後の頁 346～359
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bib/bbz153	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 該当する

1．著者名 Liu Lizhi、Mamitsuka Hiroshi、Zhu Shanfeng	4．巻 37
2．論文標題 HPOFiller: identifying missing protein?phenotype associations by graph convolutional network	5．発行年 2021年
3．雑誌名 Bioinformatics	6．最初と最後の頁 3328～3336
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bioinformatics/btab224	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 該当する

1．著者名 You Ronghui、Yao Shuwei、Mamitsuka Hiroshi、Zhu Shanfeng	4．巻 37
2．論文標題 DeepGraphGO: graph neural network for large-scale, multispecies protein function prediction	5．発行年 2021年
3．雑誌名 Bioinformatics	6．最初と最後の頁 i262～i271
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bioinformatics/btab270	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 該当する

1．著者名 Nguyen Dai Hai、Nguyen Canh Hao、Mamitsuka Hiroshi	4．巻 110
2．論文標題 Learning subtree pattern importance for Weisfeiler-Lehman based graph kernels	5．発行年 2021年
3．雑誌名 Machine Learning	6．最初と最後の頁 1585-1607
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s10994-021-05991-y	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 Guvenc Paltun Betul、Kaski Samuel、Mamitsuka Hiroshi	4．巻 22
2．論文標題 Machine learning approaches for drug combination therapies	5．発行年 2021年
3．雑誌名 Briefings in Bioinformatics	6．最初と最後の頁 bbab293
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bib/bbab293	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 該当する

1．著者名 Liao Zhirui、Huang Xiaodi、Mamitsuka Hiroshi、Zhu Shanfeng	4．巻 -
2．論文標題 Drug3D-DTI: Improved Drug-target Interaction Prediction by Incorporating Spatial Information of Small Molecules	5．発行年 2021年
3．雑誌名 Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2021)	6．最初と最後の頁 340-347
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/BIBM52615.2021.9669707	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 該当する

1．著者名	4．巻
Nguyen Dai Hai、Nguyen Canh Hao、Mamitsuka Hiroshi	-
2．論文標題	5．発行年
Machine Learning for Metabolic Identification	2021年
3．雑誌名	6．最初と最後の頁
Creative Complex Systems	329～350
掲載論文のDOI（デジタルオブジェクト識別子）	査読の有無
10.1007/978-981-16-4457-3_20	無
オープンアクセス	国際共著
オープンアクセスとしている（また、その予定である）	－

1．著者名	4．巻
Hiremath Santosh、Wittke Samantha、Palosuo Taru、Kaivosoja Jere、Tao Fulu、Proll Maximilian、Puttonen Eetu、Peltonen-Sainio Pirjo、Marttinen Pekka、Mamitsuka Hiroshi	16
2．論文標題	5．発行年
Crop loss identification at field parcel scale using satellite remote sensing and machine learning	2021年
3．雑誌名	6．最初と最後の頁
PLOS ONE	e0251952
掲載論文のDOI（デジタルオブジェクト識別子）	査読の有無
10.1371/journal.pone.0251952	有
オープンアクセス	国際共著
オープンアクセスとしている（また、その予定である）	該当する

1．著者名	4．巻
Liu Lizhi、Mamitsuka Hiroshi、Zhu Shanfeng	38
2．論文標題	5．発行年
HPODNets: deep graph convolutional networks for predicting human protein?phenotype associations	2021年
3．雑誌名	6．最初と最後の頁
Bioinformatics	799～808
掲載論文のDOI（デジタルオブジェクト識別子）	査読の有無
10.1093/bioinformatics/btab729	有
オープンアクセス	国際共著
オープンアクセスとしている（また、その予定である）	該当する

1．著者名	4．巻
Paltun Betul Guvenc、Kaski Samuel、Mamitsuka Hiroshi	19
2．論文標題	5．発行年
DIVERSE: Bayesian Data IntegratiVE Learning for Precise Drug ResponSE Prediction	2022年
3．雑誌名	6．最初と最後の頁
IEEE/ACM Transactions on Computational Biology and Bioinformatics	2197～2207
掲載論文のDOI（デジタルオブジェクト識別子）	査読の有無
10.1109/TCBB.2021.3065535	無
オープンアクセス	国際共著
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	－

1．著者名 You Ronghui、Qu Wei、Mamitsuka Hiroshi、Zhu Shanfeng	4．巻 38
2．論文標題 DeepMHCII: a novel binding core-aware deep interaction model for accurate MHC-II peptide binding affinity prediction	5．発行年 2022年
3．雑誌名 Bioinformatics	6．最初と最後の頁 i220～i228
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bioinformatics/btac225	査読の有無 無
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 Nguyen Duc Anh、Nguyen Canh Hao、Petschner Peter、Mamitsuka Hiroshi	4．巻 38
2．論文標題 SPARSE: a sparse hypergraph neural network for learning multiple types of latent combinations to accurately predict drug-drug interactions	5．発行年 2022年
3．雑誌名 Bioinformatics	6．最初と最後の頁 i333～i341
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bioinformatics/btac250	査読の有無 無
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 Liao Zhirui、Xie Lei、Mamitsuka Hiroshi、Zhu Shanfeng	4．巻 39
2．論文標題 Sc2Mol: a scaffold-based two-step molecule generator with variational autoencoder and transformer	5．発行年 2022年
3．雑誌名 Bioinformatics	6．最初と最後の頁 btac814
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/bioinformatics/btac814	査読の有無 無
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0件

〔図書〕　計0件

〔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

7．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関