

令和 3 年 5 月 19 日現在

機関番号：12301

研究種目：基盤研究(C)（一般）

研究期間：2018～2020

課題番号：18K11457

研究課題名（和文）行列分解に基づく効率的なニューラルネットワーク学習法の研究

研究課題名（英文）Matrix Factorization-based Efficient Learning for Neural Networks

研究代表者

林 克彦（Hayashi, Katsuhiko）

群馬大学・社会情報学部・准教授

研究者番号：50725794

交付決定額（研究期間全体）：（直接経費） 2,600,000円

研究成果の概要（和文）：行列分解、及び、それを多次元データに一般化したテンソル分解に基づいてニューラルネットワーク計算を効率化する手法について理論的・実験的な研究を行った。特に、ユニタリ対角化に基づいて、行列をベクトル化する手法を開発した。また、グラフ上のパスクエリによる質問応答問題をベンチマークとして、この手法の有効性を検証した。さらに、ニューラルネットワークモデルの圧縮を行うため、パラメータを二値化する手法についても開発を行った。

研究成果の学術的意義や社会的意義

ニューラルネットワークを基礎とした表現学習や深層学習は人工知能の中核を担う技術である。その計算効率の向上は様々な人工知能アプリケーションの実用化に向けて重要な意義を持つ。ユニタリ対角化に基づく手法は行列パラメータをベクトル化する汎用的な手法であり、自然言語処理や情報検索など様々な分野におけるアプリケーションの処理効率を向上させる。また、二値化についてもシステムのメモリ消費を大幅に削減し、上記分野の様々なアプリケーションの実用化に貢献するものである。

研究成果の概要（英文）：A theoretical and empirical investigation was carried out on efficient neural network computations based on matrix factorization and its generalization to multi-dimensional data, which is called tensor factorization. Especially, we developed a unitary diagonalization method to transform matrices into vectors. To show the effectiveness of the method, we conducted experiments on graph-path query answering task. In addition, we also studied a binarization method for neural network model compression.

研究分野：知能情報学

キーワード：行列分解 ニューラルネットワーク 知識グラフ

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、研究者個人に帰属します。

1. 研究開始当初の背景

近年、ニューラルネットワークを基礎とした表現学習・深層学習に関する研究が盛んに行われている。また、自然言語処理、音声処理、画像処理、ロボティクス、ソーシャル科学、生命科学など多岐に渡る分野の様々なタスクの性能を大幅に向上させるニューラルネットワークモデルが開発されている。これらニューラルネットワーク技術において、行列やテンソルによる写像操作はその根幹となる要素技術の 1 つであり、モデルの性能や計算・メモリ効率に大きな影響を与えている。写像操作の軽量化・効率化を行うことができれば、機械学習分野での基礎研究に留まらず、上記の様々な分野の進展に貢献するものと考えられ、研究的に重要な意義を持つ。

2. 研究の目的

ニューラルネットワークに基づく表現学習・深層学習の効率化に向けて、行列やテンソルによる写像操作を軽量化・効率化する技術を開発する。このような技術は世界的にも様々な研究が行われているが、基礎研究の視点から見た本研究の独自性は、行列への代数的な制限に基づいて、これを検討・実現するところにある。また、様々な分野におけるアプリケーションの実用化に貢献することも目的となる。

3. 研究の方法

行列・テンソル分解に基づく写像操作の効率化を行う。特に、問題に性質に合わせた制約を行列・テンソルに課すことで、予測時だけでなく、モデル学習時にも分解下で効率的な計算を行う方法を検討する。これにより、大規模なデータ上でも学習を行うことができる。提案するモデルの実験検証は、自然言語処理や知識グラフ処理に関するデータセットを用いて行う。

4. 研究成果

本研究課題における研究成果は大きく分けて以下の 2 つである。

(1) ユニタリ対角化に基づく行列分解

行列（やテンソルのスライス行列）を分解するため、問題に性質に合わせた制約を行列に課すことでその軽量化・計算効率化を行った。特に、正規性・可換性の制約を課すことで、行列のユニタリ対角化を行う手法を提案した。この手法を知識グラフ埋め込み、再帰型ニューラルネットワークに組み込む方法を考案し、新たな表現学習・深層学習モデルを提案した。知識グラフ埋め込みモデルは国際会議 AAAI18、EMNLP19 に採録され、再帰型ニューラルネットワークモデルは国際会議 NAACL18 に採録された。

本手法のさらなる改善に向けて、有限群論に基づく制約方法についても検討を行い、一定の成果が得られている。また、(2)で述べる二値化についても、モデルのパラメータ要素を有限要素に限定するもので、本検討の特殊なケースとして考えられる。これら成果を次のプロジェクトへ引き継ぐことも検討している。

主な研究成果のまとめとして、国際会議 NAACL18、AAAI18、EMNLP19 の各 Long Paper に採録、国内研究会・学会での優秀賞受賞、などの成果が得られた。

(2) パラメータの二値化によるモデル圧縮

行列やベクトルの各要素は実数値で表されることが一般的である。この実数値を $-1, +1$ のような二値に限定することで、モデルのパラメータ保持にかかるメモリ削減や計算効率の改善を行った。モデルの学習方法は Hinton's straight through 推定と呼ばれる手法を利用することで実数値に限定されたパラメータでも確率的勾配降下法のような勾配法でも最適化することを実現した。この手法を知識グラフ埋め込みに組み込む方法を考案し、モデル性能を落とすことなく、メモリ消費量の大幅な削減に成功した。

本手法のさらなる改善に向けて、二値を $0, 1$ に限定する手法についても検討を行った。この意図は、確率関係データベースのモデル化において、パラメータを確率値 $[0, 1]$ に限定することで、SQL のような高度な情報検索を効率的に行う方法を実現できることに気がついたためである。一方、 $0, 1$ では表現に負を利用することができないため、モデルパラメータの学習調整が難しく、現時点では、外部に報告できるような成果とはなっていない。

国際論文誌 IEEE TKDE、国際会議 ECIR19 に採録、国内学会での優秀賞受賞、などの成果が得ら

れた。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 Katsuhiko Hayashi, Koki Kishimoto, Masashi Shimbo	4. 巻 --
2. 論文標題 A Greedy Bit-flip Training Algorithm for Binarized Knowledge Graph Embeddings	5. 発行年 2020年
3. 雑誌名 Findings of the Association for Computational Linguistics	6. 最初と最後の頁 109,114
掲載論文のDOI（デジタルオブジェクト識別子） 10.18653/v1/2020.findings-emnlp.10	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 Katsuhiko Hayashi, Koki Kishimoto, Masashi Shimbo	4. 巻 --
2. 論文標題 Binarized Embeddings for Fast, Space-Efficient Knowledge Graph Completion	5. 発行年 2021年
3. 雑誌名 IEEE Transactions on Knowledge and Data Engineering	6. 最初と最後の頁 1,13
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TKDE.2021.3075070	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 -

〔学会発表〕 計7件（うち招待講演 2件/うち国際学会 2件）

1. 発表者名 Kishimoto Koki, Hayashi Katsuhiko, Akai Genki, Shimbo Masashi, Komatani Kazunori
2. 発表標題 Binarized Knowledge Graph Embeddings
3. 学会等名 41st European Conference on Information Retrieval
4. 発表年 2019年

1. 発表者名 Hayashi Katsuhiko, Shimbo Masashi
2. 発表標題 A Non-commutative Bilinear Model for Answering Path Queries in Knowledge Graphs
3. 学会等名 the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing
4. 発表年 2019年

1. 発表者名 林克彦, 真鍋陽俊, 石原敬大, 新保仁
2. 発表標題 Block H0IE: 関係行列の同時対角化に基づく知識グラフ埋め込みの問題点とその解決
3. 学会等名 研究報告音声言語情報処理 (SLP)
4. 発表年 2018年 ~ 2019年

1. 発表者名 Takahiro Ishihara, Katsuhiko Hayashi, Hitoshi Manabe, Masashi Shimbo
2. 発表標題 Neural Tensor Networks with Diagonal Slice Matrices
3. 学会等名 The 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics (国際学会)
4. 発表年 2018年 ~ 2019年

1. 発表者名 Hidetaka Kamigaito, Katsuhiko Hayashi, Tsutomu Hirao, Masaaki Nagata
2. 発表標題 Higher-order Syntactic Attention Network for Long Sentence Compression
3. 学会等名 The 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics (国際学会)
4. 発表年 2018年 ~ 2019年

1. 発表者名 林克彦
2. 発表標題 知識グラフと分散表現
3. 学会等名 第21回情報論的学習理論ワークショップ (招待講演)
4. 発表年 2018年 ~ 2019年

1. 発表者名 林克彦
2. 発表標題 知識グラフと分散表現
3. 学会等名 言語処理学会第25回年次大会（招待講演）
4. 発表年 2018年～2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

<p>大阪大学 RESOU https://resou.osaka-u.ac.jp/en/research/2019/20190522_1</p>
--

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
研究分担者	高瀬 翔 (Takase Sho) (40817483)	日本電信電話株式会社NTTコミュニケーション科学基礎研究所・協創情報研究部・リサーチアソシエイト (94305)	
研究分担者	上垣外 英剛 (Kamigaito Hidetaka) (40817649)	東京工業大学・科学技術創成研究院・助教 (12608)	

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
---------	---------