

非負行列分解の解法と応用に関する研究

メタデータ	言語: ja 出版者: 公開日: 2021-03-09 キーワード (Ja): キーワード (En): 作成者: 水谷, 友彦 メールアドレス: 所属:
URL	http://hdl.handle.net/10297/00027957

令和 2 年 6 月 15 日現在

機関番号：13801

研究種目：若手研究(B)

研究期間：2015～2019

課題番号：15K20986

研究課題名(和文)非負行列分解の解法と応用に関する研究

研究課題名(英文)Algorithms and Applications of Nonnegative Matrix Factorization

研究代表者

水谷 友彦(Mizutani, Tomohiko)

静岡大学・工学部・講師

研究者番号：00553984

交付決定額(研究期間全体)：(直接経費) 3,100,000円

研究成果の概要(和文)：分離可能性を仮定した下での非負行列分解に対する計算手法を改良、拡張し、大規模実データの分類や特徴量抽出に役立てることを研究目的とした。特に、楕円丸め法および前処理付き逐次射影法のノイズ頑強性と計算コスト削減に関して研究を実施した。その結果、前処理付き逐次射影法のノイズ頑強性に関して既存研究の結果を改善することに成功した。また、近似的に特異値分解を計算する手法を開発し、それを組み入れることで楕円丸め法や逐次射影法の計算コストを削減することに成功した。そして、ハイパースペクトラル画像の端成分抽出問題に対して有効に動作することを確認した。

研究成果の学術的意義や社会的意義

分離可能性を仮定した下での非負行列分解は新聞記事を分類する問題やハイパースペクトラル画像から端成分を抽出する問題などに応用することができる。このような実問題への応用を想定したとき、計算手法はノイズに対して頑強で、大規模データに対しても動作することが望まれる。本研究課題では逐次射影法と楕円丸め法のアルゴリズムを改良することで、ノイズ頑強性の向上と計算コストの削減に成功した。そして、実際にハイパースペクトラル画像の端成分抽出問題に適用し、その有効性を確認した。以上の考察から開発した手法は大規模実データから有益な情報を抽出することに役立つことが期待できる。

研究成果の概要(英文)：This research project aimed to enhance algorithms for computing a nonnegative matrix factorization under separability assumption, and apply to extraction of features from big data. In particular, we examined the robustness of preconditioned successive projection algorithm, and improved the results shown by previous studies. Also, we developed an algorithm for computing singular value decompositions of matrices with low cost, and incorporated it into an ellipsoidal rounding algorithm and successive projection algorithm. We applied the resulting algorithms to endmember detection problems of hyperspectral images, and confirmed that they work well even if the data size is large.

研究分野：数理最適化

キーワード：非負行列分解 特異値分解 低ランク近似 逐次射影法 体積最小閉包楕円 スペクトラル・クラスタリング ハイパースペクトル画像

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

非負行列分解(NMF)とは与えられた非負行列を二つの非負行列の積の形に分解するという問題である。NMFに関する研究は1970年代頃から行われている。特に1999年のNature誌に掲載されたLee-SeungによるNMFを利用した顔画像の基底抽出に関する研究が一つの契機となり、それ以降、活発に研究が行われている。NMFは多くの応用を持ち、例えば、文章、画像、音声から有益な情報を見つけることに利用されている。

NMFを計算することは困難である。そこで、Aroraらは分離可能と呼ばれる仮定を導入した。この仮定の下ではNMFは多項式時間で計算することができる。NMFに分離可能性を仮定すると応用の範囲は限定される。しかしながら、文章コーパスやハイパースペクトラル画像からの特徴量抽出においては、この仮定は妥当であると言われている。以降では、分離可能性を仮定した下でNMFを計算する問題をSeparable NMFと記述することにする。実問題から生じるSeparable NMFではデータ行列を厳密に二つの非負行列に分解できることは稀であり、ある程度のノイズが含まれていることを想定するのが自然である。そのため、Separable NMFに対する計算手法はデータ行列にノイズが含まれていても正しく動作することが望まれる。このような手法をノイズに対して頑強であるという。

研究代表者はSeparable NMFに対して楕円丸め法という計算手法を開発し、ノイズに対して頑強であることを理論的に示した。また、文章データのクラスタリングやトピック抽出に対して有効であるという実験結果を得た。本研究課題では、楕円丸め法の改良、拡張、大規模実データへの応用に関して研究を実施し、楕円丸め法が真に有用な手法となることを目指した。

2. 研究の目的

本研究課題では、社会から要請されるデータ分析に対して楕円丸め法が有用な手段となることを目指した。具体的には以下のような研究目標を掲げた。

- アルゴリズムの改良: 楕円丸め法が大規模データを扱えるように改良する。前処理技術を利用することでノイズ頑強性の向上を図る。
- アルゴリズムの拡張: 楕円丸め法を半正定値テンソル分解に拡張する。また、楕円丸め法とスペクトラル・クラスタリングとの類似性を調べ、楕円丸め法のクラスタリング手法への適用可能性も検討する。
- 大規模実データへの応用: 開発した手法を大規模実データへ適用し有効性を検証する。また、研究の中で開発したソースコードをソフトウェアとして整理し公開する。

3. 研究の方法

まず楕円丸め法の大規模データへの適用に関する研究方法を述べる。楕円丸め法ではデータ行列の特異値分解を計算しデータを低次元空間に射影する。そして低次元空間に射影したデータを包含する楕円の中で一番体積が小さくなるものを計算する。体積最小閉包楕円の計算は内点法と有効制約法を組み合わせると高速に計算することができる。一方で、特異値分解の計算コストは大きく、特に大規模データを扱う場合は特異値分解の計算が障害となる。そこで、特異値分解の計算精度をある程度犠牲にする代わりに計算コストを削減するという方針で研究を進めた。次に楕円丸め法のノイズ頑強性向上に関する研究方法を述べる。Gillis-VavasisはSeparable NMFに対する逐次射影法と呼ばれる計算手法のノイズ頑強性について研究を行った。その一連の研究の中で、逐次射影法のノイズ頑強性を向上させるための前処理技術を開発した。この技術は楕円丸め法に対しても有効かもしれない。楕円丸め法のノイズ頑強性を向上させる

前処理技術の開発に向けて研究を進めた。

Separable NMF に対する楕円丸め法とデータ分類法であるスペクトラル法は多くの共通点が見いだせる。スペクトラル法ではラプラシアン行列の固有ベクトルを計算し、それに対して K 平均法を適用することでデータの分類を行う。楕円丸め法との類似性を踏まえると、 K 平均法の代わりに凸最適化問題を解くという着想が自然に生まれる。 K 平均法の代わりに凸最適化を利用するスペクトラル法の性能を調査した。

ハイパースペクトル画像から端成分を抽出する問題は Separable NMF として定式化することができる。ベンチマークとしてよく利用されるハイパースペクトル画像については入手することができた。そこで、端成分抽出問題に対する楕円丸め法の性能を実験的に評価した。

4. 研究成果

アルゴリズムの改良：前処理技術の利用

Gillis-Vavasis は逐次射影法に対する前処理技術を開発しノイズ頑強性を理論的に評価した。彼らの解析ではデータの次元と分解ランクが一致するという条件が成立することを仮定している。実問題から生じる Separable NMF においてはこの条件が成立することは稀である。そこで、データの次元と分解ランクが一致するという条件の成立を仮定しないで前処理付き逐次射影法のノイズ頑強性を評価した。この研究に関して Gillis 氏とも議論を行った。研究成果をまとめた論文は Linear Algebra and its Applications 誌に掲載された。

アルゴリズムの改良：大規模データへの適用

楕円丸め法および前処理付き逐次射影法を大規模データに適用するとき特異値分解の計算が障害となる。Rokhlin らによって開発された乱択部分空間法は大規模行列に対する特異値分解に対して有効であることが知られている。この手法ではガウス行列を利用することで行列の像空間の低ランク構造を見つける。逐次射影法は列ピボット付き QR 分解と本質的に同じである。そこで、本研究ではガウス行列の代わりに逐次射影法を利用することを提案し、近似精度の理論解析を行った。また、数値実験で大規模行列に対する有効性を検証した。行列の特異値分解には Matlab の svds コマンドがよく利用される。実験では提案手法を利用すると svds コマンドとほぼ同じ精度の計算結果が非常に短い計算時間で得られることを確認した。そして、提案手法を楕円丸め法および逐次射影法に組み入れ、そのノイズ頑強性は元の手法とほぼ同程度であることを確認した。実際にハイパースペクトル画像から端成分を抽出する問題に適用し手法の有効性を確認した。研究成果は Machine Learning 誌に掲載された。

アルゴリズムの拡張

Peng らは標準的なスペクトラル法、つまり、 K 平均法を利用するスペクトラル法の理論性能を明らかにした。その解析手法を手がかりに提案手法、つまり、凸最適化を利用するスペクトラル法の性能解析に取り組んだ。まずは Peng らの解析技術を理解することから始めた。提案手法の性能を解析するという視点で解析技術を詳細に調べたところ、彼らの主結果を改善することに成功した。具体的には、より弱い仮定の下で近似性能を改善することができた。この研究成果は論文誌に投稿し、査読者から論文の出版に関して好意的な意見を頂いた。次に、Peng らの解析技術を活用して提案手法の理論解析を行った。その結果、well-clustered と呼ばれるグラフに対してならば最適な K 分割を得られるということを示すことができた。また実データを用いて提案手法の性能を実験的に評価した。その結果、提案手法の分類性能は標準的なスペクトラル法

の平均的な分類性能とほぼ同じであることが分かった．この研究成果を論文にまとめ論文誌に投稿した．現在査読中である．

5. 主な発表論文等

〔雑誌論文〕 計8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 3件）

1. 著者名 Mizutani Tomohiko、Tanaka Mirai	4. 巻 107
2. 論文標題 Efficient preconditioning for noisy separable nonnegative matrix factorization problems by successive projection based low-rank approximations	5. 発行年 2017年
3. 雑誌名 Machine Learning	6. 最初と最後の頁 643 ~ 673
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s10994-017-5673-1	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 Tomohiko Mizutani	4. 巻 -
2. 論文標題 Convex Programming Based Spectral Clustering	5. 発行年 2018年
3. 雑誌名 arXiv	6. 最初と最後の頁 -
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 無
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 Tomohiko Mizutani	4. 巻 497
2. 論文標題 Robustness analysis of preconditioned successive projection algorithm for general form of separable NMF problem	5. 発行年 2016年
3. 雑誌名 Linear Algebra and its Applications	6. 最初と最後の頁 1-22
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.laa.2016.02.016	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 Tomohiko Mizutani	4. 巻 -
2. 論文標題 Improved Analysis of Spectral Algorithm for Clustering	5. 発行年 2019年
3. 雑誌名 arXiv:1912.02997	6. 最初と最後の頁 -
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 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9件（うち招待講演 0件 / うち国際学会 2件）

1. 発表者名 水谷友彦
2. 発表標題 凸計画に基づくスペクトラル・クラスタリング手法
3. 学会等名 科学研究費・基盤研究(A)「新時代の最適化モデルに基づく意思決定支援プラットフォームの研究と開発」による2018年度ワークショップ
4. 発表年 2018年

1. 発表者名 Tomohiko Mizutani, Mirai Tanaka
2. 発表標題 Efficient preconditioning for noisy separable nonnegative matrix factorization problems by successive projection based low-rank approximations
3. 学会等名 The 9th Asian Conference on Machine Learning (国際学会)
4. 発表年 2017年

1. 発表者名 水谷友彦, 田中未来
2. 発表標題 逐次射影法と部分空間反復法に基づく低ランク近似行列計算
3. 学会等名 IBIS2016
4. 発表年 2016年

1. 発表者名 Tomohiko Mizutani
2. 発表標題 Spectral clustering by ellipsoid and its connection to separable nonnegative matrix factorization
3. 学会等名 International Symposium on Optimization (ISMP) (国際学会)
4. 発表年 2015年

〔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
--	---------------------------	-----------------------	----