



HOKKAIDO UNIVERSITY

Title	A Study on Efficient Algorithms for Nonnegative Matrix/Tensor Factorization [an abstract of dissertation and a summary of dissertation review]
Author(s)	木村, 圭吾
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第12623号
Issue Date	2017-03-23
Doc URL	https://hdl.handle.net/2115/65376
Rights(URL)	https://creativecommons.org/licenses/by-nc-sa/2.1/jp/
Type	doctoral thesis
File Information	Keigo_Kimura_abstract.pdf, 論文内容の要旨



学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称 博士（情報科学） 氏名 木村 圭吾

学 位 論 文 題 名

A Study on Efficient Algorithms for Nonnegative Matrix/Tensor Factorization
(非負値行列/テンソル分解の効率のよいアルゴリズムに関する研究)

本論文では、非負値行列分解 (Nonnegative Matrix Factorization: NMF) と非負値テンソル分解 (Nonnegative Tensor Factorization: NTF) の効率の良いアルゴリズムの提案および解析を行っている。

NMF は非負の行列を二つの非負の行列の積に分解する手法であり近年様々な分野で利用されている。従来の特異値分解 (SVD) や、主成分分析 (PCA) とは異なり、NMF による分解は比較的疎な行列が得られることが知られている。また、成分が非負であるために分解の物理的解釈が容易となり、探索的データマイニングなどにおいて有用である。しかし、NMF は SVD や PCA と異なり大域最適解を見つけることが困難であり、そのため交互最適化による局所最適解を見つける手法が数多く提案されている。ただし、交互最適化は収束するまでに多くの反復を必要とするため大規模な行列の分解には大きな計算量を必要とする。そこで反復をあまり必要としない、すなわち高速に収束する NMF アルゴリズムの開発の期待は様々な分野で高まっている。

一方、NTF は非負のテンソルを非負のテンソルと複数の非負の行列との積に分解する手法であり、NMF のテンソルへの一般化である。ここでテンソルとは多次元配列のことであり、近年センサーの技術向上やデータ収集技術及びデータの蓄積量の向上によりテンソルによるデータ表現は重要性和有用性を増している。テンソル分解は様々なテンソルデータの解析に扱われている。NTF は非負性約を持つテンソル分解であり、非負性による利点は NMF と同様に分解結果の物理的解釈可能性などが挙げられる。そのため、Recommendation といった人間の意思決定の補助となる問題では非常によく扱われる手法であり、様々な研究が行われている。行列に対する行列分解が特徴抽出や次元縮約に用いられるのと同様にテンソル分解はテンソルの特徴抽出などに有用である。テンソル分解ではテンソルを複数の行列に変換、復元することで行列分解におけるデータマイニング技術や機械学習技術を利用することが可能である。これによってテンソルデータの解析が比較的容易となる。そのため、近年テンソル及びテンソル分解を用いた手法が数多く提案されている。

NTF も NMF 同様かそれ以上に収束するまでの反復回数が問題となる。原理的には NTF の各行列、及びテンソルの最適化は NMF の最適化と等しい問題を解くことになる。そのため、収束の速い NMF アルゴリズムが適用できる。しかし、NTF の最適化におけるより大きな問題は勾配の計算にある。NTF の勾配の計算には非常に大きな行列の計算を行う必要がある。これは Intermediate Data Explosion (IDE) と呼ばれる問題であり、大規模な行列同士の積が非常に大きな計算量を必要とするために分解可能なテンソルの大きさが制限される。これに対して、分割と並列計算による高速化などが提案されているが、現在広く用いられている計算機では大規模なテンソルを分解することは難しい。

本論文では NMF については収束の速いアルゴリズムの提案と解析を行い、NTF に対しては IDE 問題を解決する効率の良い手法の提案と解析を行った。本論文の貢献は以下の通りである。

(1) 直交制約を付与した NMF である Orthogonal NMF に対して効率の良いアルゴリズムを提案した. 今まででは直交制約が行列形でしか書けない為に収束の遅い手法の適用しかできなかった. 本論文は非負制約を明示的に用いることで直交制約を変形し, 高速な手法の適用を可能にした.

(2) Orthogonal NMF で一般に用いられる近似尺度である二乗誤差から, より一般的な距離のクラスである Bregman Divergence へと拡張した. Bregman Divergence を用いた Orthogonal NMF の提案は世界初である.

(3) この拡張された手法が Bregman Hard Clustering と呼ばれる一般的なクラスタリングのクラスと等価であることを証明した. この等価性を示したことにより, 提案法は K-means だけでなく, Information Theoretic co-clustering といった様々なクラスタリング手法との等価性が示された.

(4) NTF に対して, 大規模テンソルに対して高速かつ省メモリで分解を行うスパース NTF 手法を提案した. 前述にあるように, IDE の問題を予測により予め不要な計算を行わないことで解決した. 従来のサンプリングを用いた高速化手法などは近似誤差が大きくなる恐れがあったが提案法は従来の NTF と同じ近似誤差を保証する. あるデータセットでは 5 倍以上の高速計算と同時に 1/500 程のメモリ使用量での分解を達成した.

本論文の構成は以下の通りである. まず第 1 章で研究の背景, 目的, 貢献を述べた. 次に第 2 章で論文中に用いる表記とテンソル分解に必要な演算を定義した. 第 3 章では, NMF と NTF のアルゴリズムや拡張方法, 応用についての紹介を行った. 第 4 章で貢献 (1)-(3) にあたる, 直交性を考慮した NMF に対する効率の良いアルゴリズムの提案し, その効果を示した. 第 5 章では貢献 (4) に対して, 2 種類の効率のよい NTF アルゴリズムを提案し, 実験とその効果の解析, 考察を行った. 最後に第 6 章で結論を述べた.