論文 / 著書情報
Article / Book Information

題目(和文)	非負値行列分解の統計的学習理論と複合データ分析
Title(English)	Statistical Learning Theory of Nonnegative Matrix Factorization and Multiple Data Analysis
著者(和文)	幸島匡宏
Author(English)	Masahiro Kohjima
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第11064号, 授与年月日:2019年3月26日, 学位の種別:課程博士, 審査員:渡邊 澄夫,樺島 祥介,金森 敬文,山下 真,中野 張
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第11064号, Conferred date:2019/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	数理・計算科学　系 数理・計算科学　コース	申請学位（専攻分野）： Academic Degree Requested	博士（理学） Doctor of
学生氏名： Student's Name	幸島 匡宏	指導教員（主）： Academic Supervisor(main)	渡邊 澄夫
		指導教員（副）： Academic Supervisor(sub)	

要旨（和文 2000 字程度）
Thesis Summary（approx.2000 Japanese Characters ）

E コマースやソーシャルネットワーキングサービスの普及，センサデバイスの発達等によって，ヒトやモノに関する様々な種類のデータが大量に収集・蓄積されるようになった．データを保有する組織や企業は，収集したユーザのサービス利用情報等を分析することでサービスの改善を図っており，データ分析は企業の強みと直結する重要技術と広く認識されている．人工知能(AI: Artificial Intelligence)やロボットなど，データ分析に関連する技術によって様々な仕事が自動化されていくという予測もされており，今後データ分析の重要性はさらに高まっていくと期待される．

近年のデータ分析で解析対象となるデータの多くは，非負の値を要素に持つ行列，非負値行列として表現することができることが知られている．例えば，文書の集合は行と列がそれぞれ文書と単語に対応する，各文書中における単語の出現回数を表す行列で表現できる．同様に購買の履歴も，行と列がユーザと商品，要素がユーザの商品購入数を表す行列として表現可能である．

非負値行列分解（NMF: Nonnegative Matrix Factorization)は，このような非負値行列として表現されるデータを入力とし，欠損値の補完やデータ中に潜むパターンを自動で抽出する手法である．具体的には，NMF の適用によって入力となる行列は行列の積の形へ分解される．得られた分解結果を利用することで，データ中のパターンの抽出と欠損値の補完が可能になる．例えば，文書の集合を表すデータへ適用することで，各文書における単語の出現回数をもとにスポーツ関連の文書やテクノロジー関連の文書などに自動で分類することができる．同様に購買データへ適用することで，チョコレート好きやコーヒー好きなどの購買のパターンを抽出し，各ユーザがどの購買パターンにどの程度基づいて購買するかを把握することが可能である．さらに NMF は行列を生成する統計モデルと見なすこともでき，一つの入力行列に対してだけではなく，複数の行列に対しても適用することができる．これにより，例えば，各月ごとに収集された購買データを表す複数の行列を分析することもできる．

NMF の研究において，データ分析技術としての NMF に注目し，モデルを拡張することで適用可能なデータを増やすこと(広汎化)と，統計モデルとしての NMF に注目し，NMF の統計的推測の理論的挙動を解明すること(理論解析)の 2 つは重要な研究の方向性であると認識されている．本稿では，この NMF の広汎化と理論解析に関する 2 つの研究成果を示す．

広汎化に関する 1 つ目の研究成果は，入力となる行列の行または列の粒度が異なる場合であっても適用可能な NMF の発展版モデルの構築とアルゴリズムの導出である．近年のデータ分析では，データを網羅的に収集する困難さや個人情報保護等の観点から，粒度の異なる複数のデータ，例えば``あるユーザがあるお店を何回訪問したか″というユーザ個人のデータと``ある同一年代のユーザ集団にある商品が計何個購入された″というユーザ集団のデータの組を分析する機会が増えている．提案手法によって個人単位・集団単位のデータの組のような粒度の異なる複数のデータを組み合わせて分析することが可能になる． これにより，例えば個人単位のデータの量が少ない場合であっても精度良く欠損値を補完することが可能になる．

理論解析に関する 2 つ目の研究成果は，ハイパーパラメタを持つ事前分布を導入したベイズ的アプローチに基づく NMF のアルゴリズム，変分ベイズ NMF(VBNMF: Variational Bayesian NMF)の理論解析である．VBNMF によって出力される行列の分解結果は，変分自由エネルギーと呼ばれる目的関数の最小値であり，ハイパーパラメタの設定値の影響を受けている．しかし，変分自由エネルギーの理論解析はこれまで行われておらず，ハイパーパラメタと分解結果の関係は解明されてこなかった．そこでこの研究では，変分自由エネルギーの漸近解析を行うことでこれを解明し，ハイパーパラメタの設定値によって VBNMF の分解結果が大きく変化する相転移構造が存在することを示した．この結果はハイパーパラメタ設定の指針としても利用できる．

本稿の構成は次の通りである．まず第 2 章で統計モデルとしての NMF とその推定アルゴリズムについて示す．次に第 3 章で粒度が異なる複数の行列として表現されるデータを分析する NMF の発展版モデルについて示す．第 4 章では変文ベイズ NMF の理論解析の結果を示す．最後に第 5 章でまとめる．

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。
Note：Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).
注意：論文要旨は、東工大リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。
Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	数理・計算科学 系 数理・計算科学 コース	申請学位（専攻分野）： Academic Degree Requested	博士 (理学) Doctor of
学生氏名： Student's Name	幸島 匡宏	指導教員（主）： Academic Supervisor(main)	渡邊 澄夫
		指導教員（副）： Academic Supervisor(sub)	

要旨（英文 300 語程度）
Thesis Summary（approx.300 English Words ）

A method of decomposing a nonnegative matrix into a product of low rank nonnegative matrices, nonnegative matrix factorization (NMF), is widely applied to pattern recognition and data analysis. However, mathematical theory of statistical learning and application to multiple data have not been established. In this thesis, we clarify the accuracy of statistical inference of NMF using variational Bayes method (VBNMF) and construct a method applicable to multiple data with different granularity.

Variational Bayesian method is a representative algorithm for NMF. The factorization result output by the VBNMF is determined by the contribution of the hyperparameter to the variational free energy (VFE), which is the objective function of VBNMF. However, theoretical property of VFE has not been clarified. This study investigates the property by asymptotic analysis and clarifies the phase transition structure, which describes the relation between hyperparameter and factorization result.

Due to e.g., the difficulty of comprehensive data collection and protection of personal information, it is required to analyze the data with different granularity, for example, user individual's data representing such as a visit count by user and user group's data representing such as purchase count by gender/age. Since standard NMF cannot be applied in this setting, we construct an extended NMF that can handle multiple matrices whose granularity of the rows or columns are different.

This thesis is organized as follows:
Chapter 2 introduces the model and algorithms of NMF. Chapter 3 is devoted to the proposed NMF models for analyzing multiple dataset with inconsistent granularity. Chapter4 provides our theoretical result of NMF. Chapter 5 summarizes the thesis.

備考 ： 論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。
Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

注意：論文要旨は、東工大リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。
Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).