JAIST Repository
https://dspace.jaist.ac.jp/

Title	非負数マトリックス因数分解へのリッチモデルと高速アルゴリズム
Author(s)	Nguyen, Duy Khuong
Citation	
Issue Date	2016-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/13512
Rights	
Description	Supervisor:Ho Bao Tu, 知識科学研究科, 博士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

Abstract

Experiments, observations, and numerical simulations in science with the major support of modern computers and sensor technologies are generating terabytes and petabytes of data. These datasets require rich models and fast algorithms to analyze large datasets to discover inside hidden knowledge for creating major breakthroughs in science, technology, industry, services and among others.

Nonnegative matrix factorization (NMF) is a linear powerful technique for dimension reduction, extracting latent factors and learning part-based representation. Specially, it reduces dimension of data making learning algorithms faster and more effective as they often work less effectively due to the curse of dimensionality. Moreover, latent factor extraction and part-based learning representation give concise interpretability of datasets to discover hidden knowledge. Although NMF and its applications have been developed for more than a decade, they have still several limitations of modeling and performance.

In this study, we designs rich models and fast algorithms for nonnegative matrix factorization. Specially, rich models provides concise interpretability describing data and enhance the quality of models by adding constraints to adapt the complexity of growing large datasets. In addition, fast algorithms are essential to find out these rich models for large datasets.

In summary, this study has the following contributions:

Firstly, concerning about rich NMF models, we propose a new rich NMF model as simplicial nonnegative matrix factorization and nonnegative matrix factorization with L_1 L_2 regularizations. Simplicial nonnegative matrix factorization can enhance smoothness and sparsity, and give more concise interpretability of the role of latent components over data instances. In addition, we generalize another rich NMF model as nonnegative matrix factorization with L_1 L_2 regularizations for Frobenius norm and KL divergence, which can enhance smoothness and sparsity of NMF models.

Secondly, we propose a fast parallel and distributed algorithm using limited internal memory for nonnegative matrix factorization with Frobenius norm with L_1 L_2 regularizations, which is based on the the accelerated anti-lopsided algorithm for nonnegative least squares. The proposed algorithm has fast over-bounded guaranteed convergence $O(\left[\left(1-\frac{\mu}{L}\right)\left(1-\frac{\mu}{rL}\right)^{2r}\right]^{k})$ in the space of passive variables, where convex parameter μ and Lipschitz constant L are bounded as $\frac{1}{2} \leq \mu \leq L \leq r$.

Thirdly, we propose a fast parallel randomized algorithm for NMF nonnegative matrix factorization with L_1 L_2 regularizations and *KL* divergence for large sparse datasets. The proposed algorithm has fast convergence, and utilize the sparse properties of data, model and representation. In addition, the experiments indicate that sparse models and sparse

representation are archived for large sparse datasets, which is a significant milestone in this research problem.

Fourthly, we propose a fast parallel algorithm for simplicial nonnegative matrix factorization with Frobenius norm. The proposed algorithm has guaranteed instance inference with sub-linear convergence $O(1/k)$, low iteration complexity, and easy sparsity control.

Finally, we propose a fast parallel algorithm for simplicial nonnegative matrix factorization with Kullback–Leibler divergence. The proposed algorithm has guaranteed instance inference with sub-linear convergence $O(1/k)$, and easy sparsity control. The experiments indicate that this approach can achieve highly sparse representation with higher accuracy in comparison with equivalent approaches.

In summary, this thesis discusses two significant mutual aspects of nonnegative matrix factorization as rich models and fast algorithms. Specifically, we propose rich models and their four fast parallel algorithms for nonnegative matrix factorization for two divergences, which can adapt with large scale applications and various datasets.

Keywords: Rich models, fast algorithms, nonnegative matrix factorization, parallel and distributed, Frobenius norm, KL divergence