

## 論文内容の要旨

博士論文題目

Generalization of Tensor Factorization and Applications

(テンソル因子化法の一般化とその応用)

氏名 林 浩平

A multi-dimensional array or a tensor is a common representation of relational data such as the WWW network and protein-protein interactions. Given a data tensor, tensor factorization finds low-dimensional features of it that are useful for the data analysis.

As the first contribution of this thesis, we study a tensor factorization model for a heterogeneously attributed data tensor such as that contains mixed discrete and continuous variables. The model can manage the heterogeneity by employing individual exponential-family distributions for each attribute of the tensor. The assumption of heterogeneity makes a Bayesian inference intractable, and we cast the EM algorithm approximated by the Laplace method and Gaussian process. We also extended the algorithm for online learning.

Next, we discuss a kernel-based framework for completion of a partially observed tensor. We use the observed elements as inputs for a kernel function. An efficient conjugate-gradient-based algorithm for prediction is developed, and it enables us to capture the given tensor on high (possibly infinite) dimensions with moderate computational cost. We also show that our framework includes the K-nearest neighbor and the GP-based models as special cases.

We apply the methods to real data tensors and demonstrate their performances with comparing conventional methods.

氏名	林 浩平
----	------

(論文審査結果の要旨)

情報爆発と言われる現在、情報を効率的に処理する数理アルゴリズムは技術的にも経済的にも重要な役割を果たしている。これらの情報の多くは構造を持っており、これまでは行列表現を利用した手法が多く用いられてきた。しかし実際には、本質的に3次元以上の列、すなわちテンソルで表すのが自然である情報も多い。本研究はこのようなデータを扱うためのテンソル分解法を拡張して新たな方法を提案し、その有効性を示したものである。

一つめの提案法はヘテロなデータを扱うためのテンソル分解法である。元来のテンソル解析法はヒューリスティックであり、近年開発された確率モデルに基づく手法もすべてのデータに同一の確率分布を仮定していた。本研究で提案された方法はこれを任意の指数分布族に拡張しており、たとえば正規分布とポアソン分布、あるいは二値を取る確率変数などの混在が可能となった。この場合、数学的には事後分布の計算が困難となるが、効率的に近似解を求める方法も組み込まれており、実用性は高い。

二つめはテンソルデータの欠損値の推定に関するものである。従来は低いランクを持つテンソルに分解できると仮定して欠損値推定が行われていたが、本手法ではカーネル法を応用し、観測されたデータをカーネル関数の入力として扱うことで、必ずしも低いランクのテンソルで表されないデータについても、高い推定能力を発揮する方法を提案した。

以上をまとめると、本論文は、構造を持つデータを扱う方法について、テンソル分解法を拡張した手法を提案しており、かつその有効性が実データによって確認されている。これは機械学習によるデータ解析に大きく寄与するものであり、博士(工学)の学位に値するものと認められる。