

## テンソル自己組織化写像

著者	岩崎 亘
発行年	2016-06-30
学位授与番号	17104生工博甲272号
URL	<a href="http://hdl.handle.net/10228/5689">http://hdl.handle.net/10228/5689</a>

氏名・(本籍)	岩崎 亘 (山口県)
学位の種類	博士 (情報工学)
学位記番号	生工博甲第272号
学位授与の日付	平成28年6月30日
学位授与の条件	学位規則第4条第1項該当
学位論文題目	テンソル自己組織化写像
論文審査委員会	委員長 教授 石井 和男
	” 森江 隆
	” 古川 徹生
	” 夏目 季代久
	” 田中 啓文

## 学位論文内容の要旨

本論文は関係データをモデル化し可視化するための解析技術を提案するものである。関係データとは複数要素の組み合わせから観測されるデータであり、たとえばオンラインショップの商品評価データはユーザと商品の組み合わせ（ユーザ×商品）から得られる関係データである。近年ビッグデータが注目される中で関係データ解析の必要性が高まっており、たとえば電子メールデータ（送信者×受信者×トピック）や脳波データ（チャンネル×時間×周波数）、顔画像データ（人物×表情×カメラアングル×照明）などはすべて関係データとして表される。本論文が提案するテンソル自己組織化写像(Tensor SOM: TSOM)は、これらの関係データを非線形テンソル分解することを通して、データに内在する本質を可視化する手法である。

第一章では序論として、本研究の背景と目的および本論文の構成について述べる。

第二章では本研究の基礎となる概念や手法について述べる。まずテンソルの表記法や演算について説明し、テンソルの基本的解析手法である線形テンソル分解について概説する。次に TSOM の基盤となる自己組織化写像(Self-Organizing Map: SOM)について述べ、さらに SOM を線形基底関数モデルで表現する拡張について説明する。またベイズ推論の立場から SOM を理論的に再構築した生成位相写像(Generative Topographic Mapping: GTM)についても説明を行う。

第三章では本論文の主題である TSOM について提案する。まず TSOM が解くべき問題設定について説明したのち、TSOM のアルゴリズムについて述べる。また TSOM アルゴリズムの理論的導出も行う。さらに派生形として基底関数型 TSOM およびテンソル生成位相写像(Tensor GTM: TGTM)についても述べる。

第四章では人工データを用いて TSOM の動作を検証する。このシミュレーションにより TSOM が期待する動作を行うことを示した。さらに計算速度と計算精度についても評価を行い、基底関数型 TSOM を用いれば大規模データであっても高速に計算できることを示し

た。

第五章では TSOM を利用したさまざまな可視化法を提案する。さらに実データについて応用し、TSOM の実用性を示した。このような多様な可視化法が用いられることも TSOM の長所である。本論文では寿司および飲料の嗜好調査データに対して TSOM を用いてユーザーと商品の同時解析を行い、多くの情報が引き出せることを示した。

実データにおいて関係データは欠損値を伴っていたり、付随するデータをもっていたりする。第六章ではこのようなデータに対する TSOM の拡張について述べる。まず欠損値のある関係データに対し TSOM のアルゴリズムを拡張する。これは推薦システムの基盤技術として重要である。次に属性情報とよばれる付加情報が付随する関係データへの TSOM 拡張について述べる。これらの拡張版 TSOM は映画評価データに適用し、その有効性を示した。また電子メールや SNS などのコミュニケーションデータに適した TSOM の拡張を示し、実データに応用した。

第七章では討論として、まずテンソル解析と SOM 双方に関わりのある他手法である高階自己組織化写像(SOM<sup>2</sup>)と本手法の比較を行う。このことを通して、TSOM や SOM<sup>2</sup>を包括する理論体系を構築することを試みる。これにより TSOM おとび SOM<sup>2</sup>はより大きな枠組みのもとで一般化可能なアルゴリズムファミリーを作る可能性を示した。また、本手法の基盤である SOM との関係についても議論を行い、SOM の一般化という枠組みで TSOM の意義を述べる。

第八章では総括として本論文をまとめる。

以上、本論文では関係データを非線形テンソル分解によりモデル化し可視化する手法として TSOM を提案した。本手法は線形のテンソル解析手法にくらべて柔軟な表現能力をもっており、また多様な可視化法を利用することで多くの情報を引き出すことができる。さらに複雑な構造をもつ大規模データや欠損をもつ不完全データにも応用できるため、幅広い分野での応用が見込める。したがって本研究は知的情報処理を発展させていくうえで重要な基盤技術である。

## 学位論文審査の結果の要旨

本論文に関し、調査委員からデータ規模と計算量の関係、計算限界、特別なデータ構造を持つ場合への適用などについて質問がなされたが、いずれも著者から満足な回答が得られた。

また公聴会においても多数の出席者があり、時系列データへの適用や学習の安定性・再現性などの質問がなされたが、いずれも著者の説明によって質問者の理解が得られた。

以上により、論文調査及び最終試験の結果に基づき、審査委員会において慎重に審査した結果、本論文が博士（情報工学）の学位に十分値するものであると判断した。