

Title	高次元データ解析法の開発とそのマーケティング分析への応用
Sub Title	Development of high dimensional data analysis methods and its applications to marketing analysis
Author	鈴木, 秀男(Suzuki, Hideo)
Publisher	慶應義塾大学
Publication year	2018
Jtitle	学事振興資金研究成果実績報告書 (2017.)
JaLC DOI	
Abstract	<p>近年、インターネットの普及とITの進化により大容量かつ多様なデータが取得できるようになり、例えばマーケティングの分野において、大量で多様な購買履歴データの活用が望まれている。このような多様なデータは、大量の属性や変数を持つ高次元データとして扱うことができる。</p> <p>本研究では、オンラインかつ、多変数のアクセスログデータの解析を行う場を想定し、多変量時系列データの構造変化をオンラインで分析するための効果的な手法を提案した。具体的には、Yu et al.(2015)の逐次的アルゴリズムを Jerome et al.(2008)のgraphical lassoを用いた結合ベースのベイジアンネットワークに組み込むことでアクセスログデータを対象とした構造変化検知手法を構築した。本提案手法について、ゴルフ関連ECサイトのアクセスログデータへの適用だけでなく、構造が分かっている人工データを適用する実験も行い、計算時間、変化検出力に関して優れていることを示した。他の提案手法として、密度クラスタリング手法(DBSCAN)におけるパラメータ設定が自動化されたAutoEpsDBSCANをさらに拡張し、完全自動化したShirai-DBSCANを提案した。数値実験や実データによる検証により、Shirai-DBSCANは任意のクラスタ構造や外れ値への対応を可能とすることを示した。</p> <p>さらに、高次元データ解析の基礎・理論研究として、2つの母集団の分散共分散行列が定数倍の関係にあるという帰無仮説を検定するための統計量について、その高次元条件下での漸近的挙動を調べた。また、多母集団が想定される場合のクラスタリングを伴う回帰分析について議論した。</p> <p>In recent years, the widespread use of the Internet and the evolution of Information technology enable us to acquire large volumes of various data, where for example, it is desired to utilize a large variety of purchase history data in the field of marketing. Such various data can be handled as high dimensional data that have a large number of attributes and variables.</p> <p>In this research, we proposed an effective method for online analysis for detecting the structural change of multivariate time series data by assuming the situation where online and multivariate access log data was analyzed. Particularly, a method for detecting structural changes for access log data was constructed by incorporating the sequential algorithm (Yu et al. (2015)) into a joint-based Bayesian network with graphical lasso (Jerome et al. (2008)). As for the proposed method, in addition to the application to the access log data of the golf related EC site, an experiment was also conducted by applying to artificial data whose structure was known for the proposed method, which indicated that the proposed method performs better in terms of the computation time and the power of change detections. As for another proposed method, we expanded AutoEps DBSCAN in the density clustering method (DBSCAN) that automates parameter settings, and then proposed fully automated Shirai-DBSCAN. It was shown that Shirai-DBSCAN can deal with arbitrary cluster structures and outliers though numerical experiments and verifications by real data.</p> <p>Furthermore, as a fundamental and theoretical study of high dimensional data analysis, statistical testing for the null hypothesis that two population covariance matrices are proportional to each other was investigated under a high dimensional setting, and the asymptotic behavior of its test statistic was discussed. Regression analysis with clustering was also studied under multiple populations.</p>
Notes	
Genre	Research Paper
URL	<a href="https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=2017000002-20170351">https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=2017000002-20170351</a>

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

研究代表者	所属	理工学部	職名	教授	補助額	466	千円
	氏名	鈴木 秀男	氏名（英語）	SUZUKI, Hideo			
研究課題（日本語）							
高次元データ解析法の開発とそのマーケティング分析への応用							
研究課題（英訳）							
Development of High Dimensional Data Analysis Methods and its Applications to Marketing Analysis							
研究組織							
氏名 Name		所属・学科・職名 Affiliation, department, and position					
鈴木 秀男（SUZUKI, Hideo）		理工学部・管理工学科・教授					
松浦 峻（MATSUURA, Shun）		理工学部・管理工学科・講師					
1. 研究成果実績の概要							
<p>近年、インターネットの普及とITの進化により大容量かつ多様なデータが取得できるようになり、例えばマーケティングの分野において、大量で多様な購買履歴データの活用が望まれている。このような多様なデータは、大量の属性や変数を持つ高次元データとして扱うことができる。</p> <p>本研究では、オンラインかつ、多変数のアクセスログデータの解析を行う場面を想定し、多変量時系列データの構造変化をオンラインで分析するための効果的な手法を提案した。具体的には、Yu et al.(2015)の逐次的アルゴリズムを Jerome et al.(2008)の graphical lasso を用いた結合ベースのベイジアンネットワークに組み込むことでアクセスログデータを対象とした構造変化検知手法を構築した。本提案手法について、ゴルフ関連 EC サイトのアクセスログデータへの適用だけでなく、構造が分かっている人工データを適用する実験も行い、計算時間、変化検出力に関して優れていることを示した。他の提案手法として、密度クラスタリング手法(DBSCAN)におけるパラメータ設定が自動化された AutoEpsDBSCAN をさらに拡張し、完全自動化した Shirai-DBSCAN を提案した。数値実験や実データによる検証により、Shirai-DBSCAN は任意のクラスタ構造や外れ値への対応を可能とすることを示した。</p> <p>さらに、高次元データ解析の基礎・理論研究として、2つの母集団の分散共分散行列が定数倍の関係にあるという帰無仮説を検定するための統計量について、その高次元条件下での漸近的挙動を調べた。また、多母集団が想定される場合のクラスタリングを伴う回帰分析について議論した。</p>							
2. 研究成果実績の概要（英訳）							
<p>In recent years, the widespread use of the Internet and the evolution of Information technology enable us to acquire large volumes of various data, where for example, it is desired to utilize a large variety of purchase history data in the field of marketing. Such various data can be handled as high dimensional data that have a large number of attributes and variables.</p> <p>In this research, we proposed an effective method for online analysis for detecting the structural change of multivariate time series data by assuming the situation where online and multivariate access log data was analyzed. Particularly, a method for detecting structural changes for access log data was constructed by incorporating the sequential algorithm (Yu et al. (2015)) into a joint-based Bayesian network with graphical lasso (Jerome et al. (2008)). As for the proposed method, in addition to the application to the access log data of the golf related EC site, an experiment was also conducted by applying to artificial data whose structure was known for the proposed method, which indicated that the proposed method performs better in terms of the computation time and the power of change detections. As for another proposed method, we expanded AutoEps DBSCAN in the density clustering method (DBSCAN) that automates parameter settings, and then proposed fully automated Shirai-DBSCAN. It was shown that Shirai-DBSCAN can deal with arbitrary cluster structures and outliers though numerical experiments and verifications by real data.</p> <p>Furthermore, as a fundamental and theoretical study of high dimensional data analysis, statistical testing for the null hypothesis that two population covariance matrices are proportional to each other was investigated under a high dimensional setting, and the asymptotic behavior of its test statistic was discussed. Regression analysis with clustering was also studied under multiple populations.</p>							
3. 本研究課題に関する発表							
発表者氏名 (著者・講演者)	発表課題名 (著書名・演題)	発表学術誌名 (著書発行所・講演学会)	学術誌発行年月 (著書発行年月・講演年月)				
大野拓哉, 鈴木秀男	ベイジアンネットワークモデルの変化検知手法とマーケティング効果分析への応用	2017年度統計関連学会連合大会	2017年9月5日				
白井智鷗, 鈴木秀男	DBSCANのパラメータ設定の自動化について	2017年度統計関連学会連合大会	2017年9月5日				
山下遥, 黒木学, 松浦峻	変数選択を考慮した多変量クラスタワイズ回帰分析	日本品質管理学会第47回年次大会	2017年11月25日				