様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　３０　年　　６　月　２０　日現在

機関番号：　３２６２１
研究種目：　若手研究(B)
研究期間：　2016～2017
課題番号：　１６Ｋ１６３６１

研究課題名（和文）主要点解析法に基づいたビッグデータのスモールデータ化に関する研究

研究課題名（英文）A study on the analysis of big data with conversion to small data based on the principal points

研究代表者

山下　遥（Yamashita, Haruka）

上智大学・理工学部・助教

研究者番号：９０７５４７９７

交付決定額（研究期間全体）：（直接経費）　　2,000,000円

研究成果の概要（和文）：主要点解析法に基づくビッグデータのスモールデータ化に関する研究では、データの分析が難しいとされる、複雑な構造を持ち、かつ変数の数が膨大となっているようなデータに焦点を当てて研究を展開していった。この研究では、(1)データのクラスタリングを考慮した多変量クラスタワイズ回帰分析法、(2)3相以上の複雑な構造を有するデータの分析方法に着目し、それらの効率的かつ妥当性のあるアルゴリズムを構築した。
　さらに実際のマーケティングデータにこれらのモデルを適用し、モデルの妥当性について検証するとともに、得られた結果の可視化を考慮したモデルへと拡張した。

研究成果の概要（英文）：This study focused on the big data that has the complex structure and also the number of variables are enormous. The main proposed models are 1) Multivariate clusterwise regression model and 2) Clustering analysis method considering the three mode data, and effective algorithms were proposed. We also expand the models that can be visualized. Moreover, we applied the methods to the real-world marketing data, and show the adequecy of the application.

研究分野：　機械学習

キーワード：　主要点解析法　ビジネスアナリティクス　クラスタリング　経営工学　品質管理

１．研究開始当初の背景

近年，インターネットの普及とIT，データベース技術の進化により，企業が得られるデータは大規模化かつ多様化しており，様々な情報を含んだ「ビッグデータ」をどのように取り扱うかが企業にとっての大きな課題となっている．このようなデータ分析の課題に対して

(1)解析しやすいサイズ，かつ

(2)データが単一の母集団から発生していると考えられ，

(3)2相の単純なデータ構造が想定されていた．

よって主要点解析法をビジネスにおけるビッグデータ解析へと応用する際には，主に(1)膨大な数の変数を取り扱う必要がある，(2)データが複数の母集団から発生していると仮定して分析をする必要がある，(3)3相以上の複雑なデータ構造が存在する（例えば，顧客ごとの時系列の購買データなど）という課題が存在する．よって，この3つの大きな課題に対応する新しい分析技術を確立することで，現在，解析が困難とされているような大規模なビジネスデータの解析が容易になることが期待される．

２．研究の目的

本研究では，背景に記述した3つの課題に対応して，以下に示す3つの方法を提案し，その理論的な妥当性および実務への有効性について明らかにする．これにより，ビジネスデータの解析に関する課題に対応し得るモデルを構築することが本研究の目的である．

(1)変数の数の大きさに対応し得る高速な主要点の導出アルゴリズム

(2)複数の母集団からデータが発生していると考えられる場合の解析方法

(3)3相以上の相構造をもつデータにおける主要行列解析法

３．研究の方法

上記の(2)・(3)を実現し得るモデルを提案し，それぞれに対して効率的な解の探索アルゴリズムを提案した．具体的には，(2)については，主要点の考え方に基づき，データを分割しながらその傾向を把握する方法を提案し，(3)については，まず2相の情報を用いて主要点解析法に基づきデータをクラスタリングし，さらにその代表的な点を二つのベクトルの積で表す方法を提案した．

さらに，実データに提案したモデルを適用することによって，適用例を蓄積していくと同時に，そこで問題となったデータ解析上の問題を基に新たなモデルの提案を行った．

４．研究成果

(1) 複数の母集団からデータが発生していると考えられる場合の解析方法

各クラスターに対して主成分分析を実施して複数の主成分までを用いて超平面を構成し，各クラスターを特徴づける方法を提案した．これは主要点[参考文献①]の考え方をデータの重心からデータの主成分へと拡張したモデルとなっている．さらに，解析結果を視覚化するために，低次元射影平面を構成する手続きを与える．この結果，従来の分析手法では図1のように図示されていた3つの母集団から発生している多次元データに関しても図2のように2次元空間上にその特徴を表すことができるようになった．

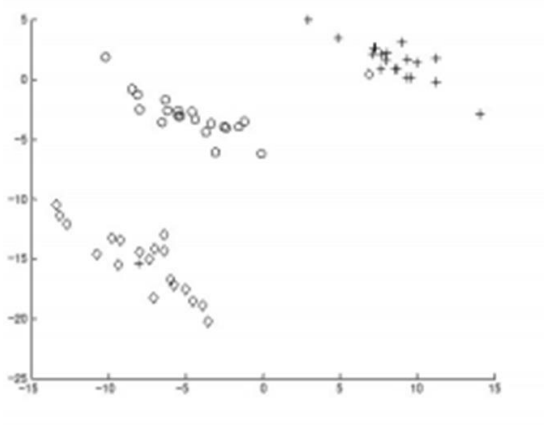

図1　k-means法に基づく
解析結果の可視化

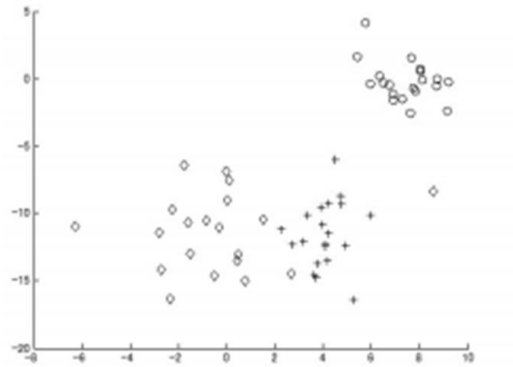

図2　提案法に基づく解析結果の可視化

本研究では，さらに解を効率的に求めるためのアルゴリズムを提案し，その妥当性についても実データ分析を通して確認した．この研究結果を

(2) 3相以上の相構造をもつデータにおける主要行列解析法

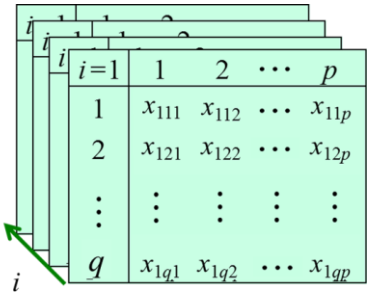

図3　3相以上の相構造をもつデータ

当該研究では，図 3 に示したような相構造を持つデータに対して，まず 1 つの相でデータを主要点解析法に基づきクラスタリングし，その後，得られたデータを図 4 のように 2 つのベクトルのクロネッカー積へと分解し，得られたベクトルの組み合わせを解釈する方法を提案した．これにより，3 相以上の複雑な構造を持つデータに対して，効率よくその内容を分析することが可能となった．

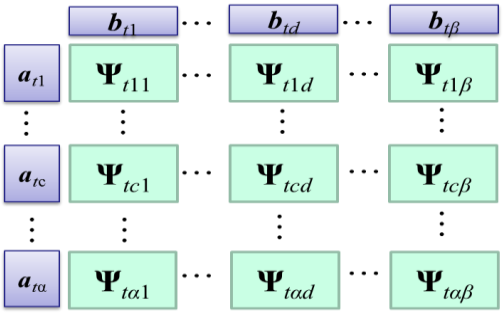

図 4　3 相以上の相構造をもつデータにおける主要点解析法

また，この問題は計算が複雑で解を求めるために膨大な計算量が必要であるといった問題が存在した．これに対して効率的なアルゴリズムを提案した．

(3)主要点分析法の実データへの応用に関する研究

ビジネス上の実データに対して主要点解析法の考え方に潜在クラスモデルの考え方を導入した新たなモデルを複数提案し，その妥当性について検討している．提案したモデルにより，各企業が抱えていたデータ解析に関する課題をそれぞれ解決している．具体的には，以下のデータに対する新たなモデルを提案している．

i) 就職ポータルサイトにおける企業と学生とのマッチングデータ
ii) 新聞における記事およびそのカテゴリデータ
iii) グルメサービスサイトにおけるレストランの推薦投稿へのリアクション数に関するデータ

上記のデータは，いずれも大きなサイズのデータとなっており，また，その構造が複雑であることから解析が難しいとされていたが，提案モデルにより，分析が容易になったことを示すことができた．

＜引用文献＞

① Haruka Yamashita, Shun Matsuura, Hideo Suzuki, “Estimation of Principal Points for a Multivariate Binary Distribution Using a Log Linear Model”, Communications in Statistics-Simulation and Computations, Vol. 46, pp. 136--147, 2017.

５．主な発表論文等

（研究代表者，研究分担者及び連携研究者には下線）

〔雑誌論文〕（計８件）

① Haruka Yamashita, Masayuki Goto, “The Analysis Based on Principal Matrix Decomposition for 3-Mode Binary Data”, Asian Journal of Management Science and Applications, Vol.3 pp. 24--37, 2017.査読有
DOI: 10.1504/AJMSA.2017.083504

② Haruka Yamashita, Shun Matsuura, Hideo Suzuki, “Estimation of Principal Points for a Multivariate Binary Distribution Using a Log Linear Model”, Communications in Statistics-Simulation and Computations, Vol. 46, pp. 136--147, 2017. 査読有
DOI: 10.1080/03610918.2014.992541

③ 黒木学, 山下遥, “改良型 k-planes クラスター分析法と解析結果の視覚化について”, 日本経営工学会論文誌, Vol. 68, pp. 1—12, 2017. (学会賞受賞) 査読有
DOI: 10.11221/jima.68.1

④ 坂元哲平, 山下遥, 後藤正幸, 荻原大陸, “就職ポータルサイトにおける企業のアピールポイントと学生の志望理由のマッチング分析モデルに関する一考察”, 情報処理学会論文誌, Vol.58, pp.1535—1548, 2017. 査読有
ISSN: 1882-7764

⑤ 鈴木 玲央奈, 山下 遥, 後藤正幸, ″同一カテゴリ内での二値判別を許容する符号表に基づく ECOC 多値判別法″, 情報処理学会論文誌, Vol.58, pp. 2046--2059, 2017. 査読有
ISSN: 1882-7764

⑥ 劉 佩潔, 山下 遥, 岩永二郎, 樽石将人, 後藤正幸：″グルメサービスにおけるレストラン推薦投稿へのリアクション数増加を目的とした潜在クラスモデル分析″, 情報処理学会論文誌, Vol.59, No.1, pp.211-226, 2018. 査読有
ISSN: 1882-7764

〔学会発表〕（計 20 件）

① Haruka Yamashita: “An Algorithm for Principal Points Considering External Criterion for Multivariate Binary Distributions”, The 18th Asia Pacific Industrial Engineering and Management System Conference (APIEMS2017), ID163, Yogyakarta, Indonesia, 2017 年 12 月 3-6 日

② Teppei Sakamoto, Haruka Yamashita, Masayuki Goto, Jiro Iwanaga: ″A Model

for Relational Analysis of Recommendation Articles and Reactions on Gourmet Service Site", The 18th Asia Pacific Industrial Engineering and Management System Conference (APIEMS2017), ID163, Yogyakarta, Indonesia, 2017 年 12 月 3-6 日

③ Haruka Yamashita: A study on a purchasing data analysis method considering the customer", 15th Asian Network for Quality Conference (ANQ2017), Soaltee Crowne Plaza, Kathmandu, Nepal, NKT-05, 2017 年 9 月 20-21 日.

④ Yuri Nishio, Hiroaki Itou, Haruka Yamashita, Masayuki Goto: "A New Analytical Model for Customer Growth Considering Potential Purchasing Preferences", 15th Asian Network for Quality Conference (ANQ2017), Soaltee Crowne Plaza, Kathmandu, Nepal, NKT-05, 2017 年 9 月 20-21 日 (received the ANQ2017 best paper award)

⑤ Ryota Kawabe, Hiroaki Itou, Haruka Yamashita, Masayuki Goto: "Proposal of Hierarchical Structure Learning of Bayesian Network for Analyzing Customer Purchasing Behavior", 15th Asian Network for Quality Conference (ANQ2017), Soaltee Crowne Plaza, Kathmandu, Nepal, ICT-07, 2017 年 9 月 20-21 日

⑥ Ryotaro Shimizu, Teppei Sakamoto, Haruka Yamashita, Masayuki Goto: "Proposal of a purchase behavior analysis model on EC site considering questionnaire data", 15th Asian Network for Quality Conference (ANQ2017), Soaltee Crowne Plaza, Kathmandu, Nepal, SQC-12, 2017 年 9 月 20-21 日

〔図書〕（計 0 件）

〔産業財産権〕

○出願状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等
個人ホームページ
http://pweb.cc.sophia.ac.jp/yamashita_lab/

６．研究組織
(1)研究代表者
山下　遥（YAMASHITA HARUKA）
上智大学・理工学部・助教
研究者番号：　90754797

(2)研究分担者：なし
（　　）

研究者番号：

(3)連携研究者：なし
（　　）

研究者番号：

(4)研究協力者:なし
（　　）