

混合truncatedモデルと決定木とを組合せた不完全データの高精密解析法

(17510127)

平成17年度-平成19年度科学研究費補助金（基盤研究（C））研究成果報告書

平成20年3月

研究代表者 廣瀬英雄

(九州工業大学情報工学部システム創成情報工学科)

はしがき.....	ii
研究成果 要旨.....	1
研究成果	4

はしがき

研究組織

研究代表者：廣瀬英雄（九州工業大学情報工学部システム創成情報工学科）

交付決定額（配分額）

（金額単 ；千円）

	直接経費	間接経費	合計
平成17年度	700	0	700
平成18年度	600	0	600
平成19年度	500	150	650
総計	1,800	150	1,950

研究成果

(1) 学会誌等

廣瀬英雄, 鶴康太郎, 決定木とワイブルべき乗則とを組み合わせた電力ケーブルの劣化要因分析とその余寿命推定, 日本信頼性学会誌, 31, 195-203 (2008)

H. Hirose, The mixed truncated model with applications to SARS, Mathematics and Computers in Simulation, 74, 443-453 (2007)

T. Yukizane, S. Ohi, E. Miyano, H. Hirose, The bump hunting method using the genetic algorithm with the extreme-value statistics, IEICE Transactions D, E89-D, 2332-2339 (2006)

H. Hirose, More accurate breakdown voltage estimation for the new step-up test method in the gumbel distribution model, European Journal of Operational Research, 177, 406-419 (2007)

H. Hirose, A. Todoroki, Random number generation for the generalized normal distribution using the modified adaptive rejection method, Information, 8, 829-835 (2005)

F. Zaman and H. Hirose, Nice bagging and robust bagging algorithms in small sample classification, INSC2008, March 13-15, 2008, Tokyo

H. Hirose and T. Yukizane, The accuracy of the trade-off curve in the bump hunting, 7th Hawaii International Conference on Statistics, Mathematics and Related Fields, January 17-19, 2008, Honolulu

Hirose, H., Yukizane, T., and Deguchi, T., The bump hunting method and its accuracy using the genetic algorithm with application to real customer data, IEEE CIT2007, October 16-19, 2007, Aizu University

- H. Hirose, S. Ohi, and T. Yukizane, Assessment of the prediction accuracy in the bump hunting procedure, 6th Hawaii International Conference on Statistics, Mathematics and Related Fields, January 17-19, 2007, Honolulu
- H. Hirose, Estimation for the number of fragile samples in the truncated and truncated models with application to the case fatality ratio for the infectious diseases, 6th Hawaii International Conference on Statistics, Mathematics and Related Fields, January 17-19, 2007, Honolulu
- H. Hirose, T. Yukizane, E. Miyano, Boundary Detection for Bumps using the Gini's Index in Messy Classification Problems, CITSA 2006, July 20-23, 2006, Orlando, Florida
- H. Hirose, S. Matsuda, M. Hikita, Electrical Insulation Diagnosing using a New Statistical Classification Method, ICPADM2006, June 26-30, 2006, Bali, Indonesia
- H. Hirose, T. Matsumoto, Mixed Truncated Model with Application to the Lifetime Estimation of Malcondition Items, 52nd Annual Reliability and Maintainability Symposium, January 23 - 26, 2006, Newport Beach, CA
- H. Hirose, Mixed truncated model of Type II with application to the lifetime estimation of malcondition items, 5th Hawaii International Conference on Statistics, Mathematics and Related Fields, January 16-18, 2006, Honolulu
- Hirose, T. Yukizane, E. Miyano, The bump hunting method using the genetic algorithm and the extreme-value statistics with application to a messy customer database, 5th Hawaii International Conference on Statistics, Mathematics and Related Fields, January 16-18, 2006, Honolulu
- H. Hirose, Optimal boundary finding method for the bumpy regions, IFORS2005, July 11-15, 2005, Honolulu
- H. Hirose, T. Ohhata, Y. Kotou, S. Matsuda, M. Hikita, T. Nishimura, S. Ohtsuka, S. Matsumoto, S. Tsuru and J. Ichimaru, A method to classify the signals from artificially prepared defects in GIS using the decision tree method, ISEIM2005, June 5-9, 2005, Kitakyushu

研究成果 による工業所有権の出願・取得状況
なし

研究成果 要旨

研究課題名：混合truncatedモデルと決定木とを組合せた不完全データの高精密解析法

標記の研究成果を出すため、データマイニング手法と不完全データ解析法との融合をはかる研究を始める前に、基本となるデータマイニング手法、不完全データの解析法についてもそれぞれ研究を行い、基礎的な研究を進めながら両者の融合をはかった。データマイニング分野については、分類が困難なデータを分類する問題をbump hunting法を用いて解決する方法とその分類精度について研究成果を出した。その成果は、[1]の他、多くの国際会議録におさめられている。不完全データ分野については、truncated modelに関して、IEEE Reliabilityへの投稿論文[2]がIEEE Reliability Japan Chapter Award (2006)となり、受賞講演の他、日本信頼性学会への解説記事[3]が掲載された。更に、医学関連分野への応用として[4]の成果を発表した。融合研究については、実際の電力ケーブルの劣化データをもとにして、決定木と最尤推定法を用いた余寿命推定法とを組み合わせた研究成果[5]を発表することができ、初期の目的を達成することができた。

[1] T. Yukizane, S. Ohi, E. Miyano, H. Hirose, The bump hunting method using the genetic algorithm with the extreme-value statistics, IEICE Transactions D: on Information and Systems, Vol. E89-D, No. 8, pp. 2332-2339 (2006) Invited Papers from New Horizons in Computing

[2] H. Hirose: The Truncated Model and Its Applications to Lifetime Analysis: Unified Censored and Truncated Models, IEEE Transactions on Reliability, Vol. 54, No. 1, pp. 11-21 (2005)

[3] 廣瀬：トランザードデータアナリシス, 日本信頼性学会誌, Vol. 30, No. 9, pp. 337-344 (2006)

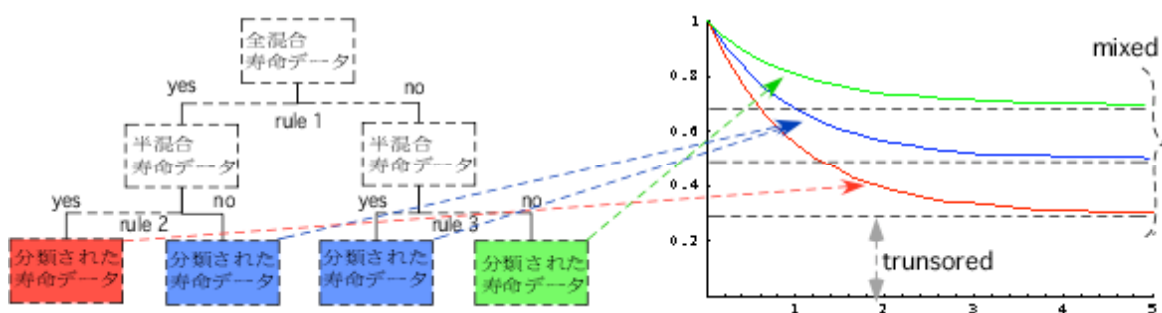
[4] H. Hirose: The mixed truncated model with applications to SARS, Mathematics and Computers in Simulation, Vol. 74, No. 6, pp. 443-453 (2007)

[5] 廣瀬, 鶴: 決定木とワイブルべき乗則とを組み合わせた電力ケーブルの劣化要因分析とその余寿命推定, 日本信頼性学会誌, Vol. 31, No. 3, pp. 195-203 (2008)

以下に示す交付申請時の研究目的に対して、平成17-19年度の実績が得られ、研究成として次のことが得られた。

研究目的（交付申請時）

不完全データの代表的な例として例えば寿命データを考える。データの背後には均質的な確率分布を仮定するのが従来の方であったが、現実的には均質的であることはまれであり、さまざまな特性をもったデータが混在している。このとき、条件付き確率（truncated data）や混合モデルを利用する方法もあるが、truncated data を使えば推定だけでなく検定問題に特に有用であった。[1]に見るように、単一の truncated data をいくつか混合させるモデルも現実問題として寿命推定には有効である。一方、寿命データはそれぞれ環境や製造法などの属性を持っており、これらを総合的に取り扱うことでより精密な（個性的な、そして信頼度の高い）寿命推定が可能になる。このことはパターン分類にデータマイニング（例えば決定木）機能や概念を与え mixed truncated model とを組み合わせることで可能である。決定木の分類基準は通常 entropy や Gini 指標であるがこれを尤度でも与え適用性を高くする。



つまり、不均質なデータを均質なデータ群に分類することと不完全データのパラメータ推定、検定問題を同時に行うことで、より現実の問題に対処した高精度の解析法を確立させる。

また、得られた数学的モデルを、工業製品の信頼性解析、医学分野の癌患者余寿命解析、流行性疫病（例えば SARS）などの現実的で重要な課題に適用し、「安全、安心」社会への寄与を目指す。

研究実績の概要（平成17年度）

不完全データを扱う新モデルとして提案しているtruncated modelを利用して、信頼性工学分野をはじめとして医学薬学分野において、寿命推定やその応用について現在までに適用を実施してきた。特に信頼性工学分野では、IEEE Reliability Japan Chapter Awardするなど、その価値がみとめられつつある。本研究課題での17年度の目標は、このtruncated modelに加えて決定木による分類を加えた精密な推定問題の解析法を構築することである。このため、まず決定木の分類問題への取り組みについて注力するため、具体的な問題を考えながら分類問題に取り組んだ。その際、比較的分類を行いやすいデータとそうでないデータとがあり、分類困難なデータの場合は未開拓分野であることから、この方面に特に注力した。分類困難なデータの場合、対象とする目的変数が他の目的変数よりも密度が濃い場所の探索 (bump hunting) 法として問題を考えることにより、確率的方法 (遺伝的アルゴリズム) と統計的方法 (極値分布による最大値の推定) とを組み合わせ、決定木を用いてbump huntingを効率的に確実にを行う方法を提示した。また、truncated modelについては、混合、truncated modelを用いて、故障データと異常状態データとの混在による新解析法を示し、またepidemiology関連としてSARS拡散モデルにも適用した。18年度は、これらの二つ (truncated modelと決定木) を融合しながら精密な解析法を構築する。

研究実績の概要（平成18年度）

不完全データを扱う新モデルとして提案しているtruncated modelを、信頼性工学分野をはじめとして医学薬学分野において、寿命推定やその関連分野について現在までに研究を行ってきた。信頼性工学分野では、IEEE Reliability Japan Chapter Awardし、その受賞講演を8月に行った。このtruncated modelに、決定木による分類を加えた精密な推定問題の解析法を構築するため、17年度から主に決定木の分類問題への取り組みについて注力し、18年度は比較的分類を行いやすいデータとそうでないデータとが混合された分類困難なデータ分類問題に対して、対象とする目的変数が他の目的変数よりも密度が濃い場所の探索 (bump hunting) 法を解決する方法について提案した。また、分類困難なデータの場合の決定木の利用については、教師データによる分類精度とテストデータによるそれとの乖離が大きいことを指摘し、それを回避する方法の提案を行った。次に、実際のフィールドデータを用いて、劣化要因を説明変数としたとき劣化の主要因を分類した後に寿命推定を行う方法について研究を開始した。一方で、truncated modelを拡張した混合truncated modelをepidemiology関連分野でのSARS拡散モデルにも適用し、その際にtruncated modelにおけるサンプル数の推定についての問題点につ

いて指摘した。19年度は、truncated modelと決定木とを融合しながら更に精密な解析法を構築する予定である。

研究実績の概要（平成19年度）

不均質なデータを、均質なデータ群から成るいくつかのグループに分類しながら、不完全データのパラメータ推定や検定問題をそれぞれのデータ群により行うことと同時に進めることで、現実の問題により対処できる高精度の解析法を確立することが目的である。この目的のために、データマイニング手法と不完全データ解析法との融合をはかる研究を始めた。不完全データを扱う新モデルとして提案しているtruncated modelを、信頼性工学分野をはじめとして医学薬学分野においても、寿命推定やその関連分野について研究範囲を広げることを行ってきた。このtruncated modelに、決定木による分類を加えた精密な推定問題の解析法を構築するため、17年度は、主に決定木の分類問題への取り組みについて注力し、18年度は、分類困難なデータ分類問題にbump hunting法を用いて解決する方法と分類精度について研究成果を出した。19年度は、実際のフィールドデータを用いて、劣化要因を説明変数としたとき劣化の主要因を分類した後に寿命推定を行う方法についてその方法論を示した。この成果を信頼性学会シンポジウムで発表し、続いてtruncated modelと決定木とを融合した精密な解析法を論文誌に発表した。一方、truncated modelを拡張した混合truncated modelを、epidemiology関連分野でのSARS拡散モデルにも適用し、その結果を論文誌に発表した。また、その際にtruncated modelにおけるサンプル数の推定についての問題点についても研究会で指摘した。これらのことから、当初の目的を十分な成果を上げることができたと考える。