様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２７ 年　 ６ 月　 ８ 日現在

機関番号：１７１０２
研究種目：基盤研究(C)
研究期間：2012～2014
課題番号：２４５０００１８

研究課題名（和文）記述長最小原理の数理と学習理論

研究課題名（英文）The Minimum Description Length Principle and Learning Theory

研究代表者
竹内　純一（Takeuchi, Junichi）

九州大学・システム情報科学研究科（研究院・教授

研究者番号：８０４３２８７１

交付決定額（研究期間全体）：（直接経費）　　4,100,000円

研究成果の概要（和文）：記述長最小原理(MDL原理)について，確率的コンプレキシティ(SC)の評価と，SCを達成する最適な予測法(minimax予測)の研究を実施した．具体的には，i.i.d.の正則なi.i.d.モデルについて，ターゲットクラスに関するJeffreys事前分布を用いたBayes混合と，その局所指数族バンドル上のBayes混合の組み合わせによってSCが達成できることを示した．特に混合型分布族および木情報源モデルについては，データ列に関する制約を除去することに成功した．また通信路容量を達成する誤り訂正符号の一つであるスパース重ね合わせ符号(Barronら，2011-)の研究を行った．

研究成果の概要（英文）：Concerning the minimum description length (MDL) principle, we performed the studies on evaluation of the stochastic complexity (SC) and the data compression and sequential prediction strategies which achieves the SC for various target models. Those algorithms performs as the minimax strategies with respect to logarithmic or coding regret. We obtained the minimax strategies when the target model is the i.i.d. general smooth families of probability distributions. Our strategy is a combination of the mixture of the target class with Jeffreys prior and a mixture of the local exponential family bundle of the target class. In particular, we succeeded to remove the restriction on the set of data sequences for the mixture family case and tree model case. We also studied sparse superposition code (Barron et al. 2011-), which is one of capacity achieving codes.

研究分野：情報理論と機械学習

キーワード：記述長最小原理　MDL　確率的コンプレキシティ　指数型分布族　木情報源モデル　局所指数族バンドル　Jeffreys事前分布

１．研究開始当初の背景

情報科学やその応用分野における機械学習技術の重要性は増し続けている．これは，計算機とインターネットの発展による必然的な帰結であろう．こうした中，情報理論に起源をもつ記述長最小原理(MDL 原理)[A2][A4][A5]は，機械学習の指導的原理の一つとして認識されている．1990 年代までには，MDL 原理の基本的考え方が確立され，応用研究が主流となった感があった．しかしながら 2006 年に，一部のモデルについて，記述長の下限である確率的コンプレキシティ(SC)を効率的に計算するアルゴリズム[A9]が発見されてから，その基礎研究が再び活発化した．

MDL 原理とは, 機械学習の過程がデータ圧縮と表裏一体であり，最小の符号長を追求することで高精度な学習が可能になるという考え方である．これを情報理論の枠組みにおいて定式化したのが J. Rissanen[A10]である．特に統計的モデル選択の文脈では，情報量規準の一つである MDL 規準がよく知られている．これは「全記述長=モデル記述長+データ記述長」が最小になるモデルを選ぶ方法である．

MDL 規準の導出自体が MDL 原理に基づいていることは重要である．すなわち，全記述長が出来るだけ小さくなるようにモデルのための符号化を最適化して得られる最小値が MDL 規準である．この考え方は，その後さらに徹底され，MDL 原理の中心概念として，確率的コンプレキシティ(Stochastic Complexity; SC)の概念が確立された[A11]．SC は，パラメトリックモデル$S=\{p(\cdot\mid\theta):\theta\in\Theta\subseteq\mathfrak{R}^d\}$を一つターゲットモデルとして固定したもとで，データの集合 W に対する regret の最悪値を最小にする符号(minimax 符号)の符号長である．ここに，minimax 符号は逐次型学習方式を導くため，SC の値を求めることは，最適な逐次学習(オンライン学習)の手法に結びつくという意味でも重要である．

こうした理由から，様々なターゲットクラスについて, SC の値を評価する努力がなされている．1990 年代には, ある正則条件のもと，SC の値が次のようになることが知られていた[A15][A12]．

$$-\log p\left(x^n\middle|\hat{\theta}\right)+\frac{d}{2}\log\frac{n}{2\pi} \qquad \text{(S1)}$$
$$+\log\int_K |J(\theta)|^{1/2}d\theta+o(1)$$

ただし，$J(\theta)$はθの Fisher 情報行列，$|J(\theta)|$はその行列式を表す．また，$o(1)$は n が大きくなるにつれてゼロに収束する量を表す．データ列 $x^n=x_1x_2...x_n$ の範囲については，最尤推定値が$K\subseteq\Theta$ に属するように制限する．

上記の評価が成り立つのは，離散無記憶情報源がなすクラスを含む指数型分布族の場合，および有限次数の Markov モデルの場合である．これらの結果の多くは Jeffreys 事前分布を用いた Bayes 符号(Jeffreys 符号)を用いて示される．このとき，離散無記憶情報源と Markov モデルの場合を除き, データ集合にやや不自然な制約を設ける必要があった．

その後, SC の評価については大きな進展は見られなかったが, 先述のように 2006 年に離散無記憶情報源の SC の値を効率的に計算するアルゴリズムが発見[A9]され, MDL 原理に関する関心が高まることとなった．

２．研究の目的

下記の項目を考察することで, MDL 原理についてより深い理解を獲得し，機械学習をはじめとするより広い分野に適用可能にすることを目指した．

・一般のパラメトリックモデルに関して SC を達成する Bayes 戦略の解析．
・Markov モデルにおける Jeffreys 符号の近似計算法の提案．
・多項 Bernoulli モデルと Markov モデルの最尤符号の効率的計算アルゴリズムの検討．
・一般の tree model に関して SC を達成する戦略の確立と解析．

ここに，Bayes 戦略とは次式で表される Bayes 混合に基づく符号を用いる手法である．

$$m(x^n)=\int p(x^n|\theta)w(\theta)d\theta$$

ただし，$w(\theta)$は事前分布と呼ばれるΘ上の確率分布である．

３．研究の方法

MDL 原理の数理的構造を解明する研究と，オンライン予測に関する応用研究を並行して行った. 平成 24 年度から 25 年度にかけては，一般の正則モデルの SC の評価を行った．また，Markov モデルの SC の近似計算法について考察し，数値シミュレーションによりその評価を行った．さらに平成 25 年度以降は，一般の tree model の SC について考察した．

また，平成 24 年度半ばからは，通信路符号化の研究を開始した．これは研究開始当初は計画していなかったテーマである．取り組んだのは，Gauss 通信路が対象であるスパース重ね合わせ符号[A7][A8]である．この符号は，Barron らが提案したもので，Shannon 限界を達成することが保証される符号の一つである．通信路符号化における復号問題は統計的モデル選択と同じ問題であることに注意すると，Shannon 限界はモデル選択の性能限界と関係することが分かる．本研究では，従来よりも現実的な設定で Shannon 限界を達成することを目指した．

４．研究成果

(1) 一般の正則モデルの確率的コンプレキシティに関する研究

SC は minimax 符号の符号長として定義される[A11]. ここでは符号とは確率分布 q と同等であり，q の regret は$\log\left(1/q(x^n)\right)-\log(1/p(x^n|\hat{\theta}))$で定義される($\hat{\theta}$はデータ x^n を与えられたもとでのθの最尤推定値). ここに, minimax 符号は，次式で表される Shtarkov の

正規化最大尤度(Normalized Maximum Likelihood; NML)に一致する．ただし，データ x の範囲が連続である場合は，総和が積分に置き換えらえる．

$$\widehat{m}(x^n) = \frac{p(x^n|\hat{\theta}(x^n)}{\sum_{x^n} p(x^n|\hat{\theta}(x^n))}$$

一般には正規化最大尤度の値を求めることは難しいが，ターゲットクラスが指数型分布族であれば，Jeffreys 事前分布を採用した Bayes 符号(Bayes 混合に基づく符号)が，漸近的に SC を達成することが知られていた[A12][A5]．指数型分布族は，確率密度関数の族が次の式で表されるモデルの種類を表し，統計学，情報理論，および機械学習において重要な役割を果たす[A1]．

$$p(x|\theta) = \exp\left(\sum_{i=1}^{d} \theta_i T(x) - \psi(\theta)\right)$$

また，Jeffreys 事前分布とは，$|J(\theta)|^{1/2}$ に比例する確率密度，すなわち

$$w_J(\theta) = \frac{|J(\theta)|^{1/2}}{\int |J(\theta)|^{1/2}\, d\theta}$$

である．この問題においては，ほとんどの場合データ集合に関する条件は厳しく，$W = \{x^n: \hat{\theta}(x^n) \in K\}$なる形に制限する必要がある．ただし，$K$ はΘの内部Θ° に包含されるコンパクト集合である．この制限は，最尤推定値が Θ の境界から一定の距離以上離れている必要があることを意味する．指数型分布族のための Jeffreys 事前分布による minimax 戦略は

$$m_n(x^n) = (1 - \epsilon_n) m_J(x^n) + \epsilon_n \int p(x^n|\theta) w_b(\theta) d\theta$$

なる形で得られる．ここに，w_b は K の境界付近で w_J よりも十分大きな値を持つ事前分布を表し，ϵ_nは適当な速さで 0 に収束する非負の数列である．この符号を修正 Jeffreys 符号と呼ぶ．第 2 項は，Jeffreys 符号 m_J の尤度が境界付近では minimax 値よりも小さくなってしまう現象に対処するために用いられる．この修正は，$K = \Theta$ の場合(すなわちデータ列を制限しない場合)は，通常用いることが出来ないが，離散無記憶情報源の族と定数次数の Markov 情報源の族の場合は可能である．

ここで，もしターゲットクラス S が指数型でない場合は，仮に最尤推定値が境界に近づかなくても，m_J の regret は minimax 値を達成しない．これは S の指数曲率が 0 でないことに由来する．

この問題を解決するため，本研究では S をより広い空間 $\bar{S}$に拡張し，$\bar{S}$上の Bayes 符号を用いる手法を提案した．すなわち，次の局所指数族バンドル[A1]を用いて S を拡張する．

$$\bar{p}(x^n|u) = p(x^n|\theta) \exp\left(\beta \cdot V(x^n|\theta) - \psi(\theta, \beta)\right)$$

$$V(x^n|\theta) = \hat{J}(x^n|\theta) - J(\theta)$$

ただし，$\hat{J}(x^n|\theta)$は，$(-1/n)\log p(x^n|\theta)$のヘシアン，すなわち経験的 Fisher 情報量を表す．Ψは正規化定数に対応する．また，β は $J(\theta)$と同じ次数の行列であり，$u = (\theta, \beta)$としている．行列 A と B に対する$A \cdot B$は要素同士の積の和を表す．このとき，漸近的 minimax 戦略として

$$\bar{m}_n(x^n) = (1 - \delta_n) m_n(x^n) + \delta_n \int \bar{p}(x^n|u) \bar{w}(u) du$$

が得られ，SC の値は(S1)となる．ただし，δ_n は適当な速さで 0 に収束する非負の数列であり，$\bar{w}$は u の適当な事前分布である．このとき，一般にはデータ系列については，$K \subseteq \Theta^\circ$ に対応する制約が必要である．また，このほか S に課される正則条件には，よく知られたモデルについても自明とは限らないものが含まれる．

このことに関して考察を進め，見通しの良い形で整備した．その結果，一般の混合型分布族が正則条件を満たすことを明らかにした．混合型分布族とは，d+1 個の確率分布 p_i (i=0,1,2,…,d)を用いて

$$p(x^n|\theta) = \sum_{i=1}^{d} \theta_i p_i(x^n) + \left(1 - \sum_{i=1}^{d} \theta_i\right) p_0(x^n)$$

で定められる確率モデルである．ただし，θ は d 次元のパラメータであり，

$$0 \le \sum_{\mathrm{i}=1}^{\mathrm{d}} \theta_i \le 1$$

を満たすものとする．これは離散無記憶情報源の族を特別な場合として含む．また，指数型分布族と双対の関係にあるクラスであり，情報幾何学において重要な役割を果たす[A1]．

混合型分布族に関しては，対象となるデータ系列に関する制限が必要ないことを，本課題において証明した(Takeuchi and Barron 2014)．この結果は，指数型分布族以外で，初めてデータに関する制約を取り除いた重要な成果である．

また，この手法をユニバーサルポートフォリオに適用する方法を考察した．ユニバーサルポートフォリオは混合型分布族に対応し，局所指数族バンドルが有効に働くケースとなる．平成 24 年度は正則化条件について調べたところ，一般には成り立たないことが明らかとなった．

(2)木情報源モデルの確率的コンプレキシティに関する研究

有限次数の Markov 情報源からなるモデルである木情報源モデル(tree model)について考察した．このモデルは，文脈木によって規定される情報源のモデルであり，データ圧縮でよく用いられる[A14]．木情報源モデルは，木の形が特別な条件を満たして，Finite State Machine (FSM)を定義する場合に限り，（データ数について）漸近的に指数型分布族に近づくことが分かっている[A13]．また，木情報源

モデルを定める木の深さを k とするとき，このモデルの任意の要素は k 次の Markov 連鎖となる．有限次数の Markov 連鎖全体は，漸近的に指数型となるため，FSM でない木情報源モデルは曲指数型分布族に近づくことを意味する．このモデルについても，混合型分布族の場合と同様の手法によって，データ列に関する制約なしに SC が達成でき，その値が (S1) となることを証明した (Takeuchi and Barron 2014)．この結果は，FSM モデル（漸近的に指数型）について知られている命題 (Takeuchi et. al 2013)の拡張でもある．

(3) スパース重ね合わせ符号の研究

当初計画にはなかった課題として，スパース重ね合わせ符号(Barron ら，2011-)の研究を実施した．これはガウス通信路に関する符号であり，辞書行列とスパースベクトルの積により符号語を構成する．このとき，辞書行列をガウス分布に従うランダム行列とすると，圧縮センシングと同等の問題となる．

この符号は，効率的復号アルゴリズムのもとで通信路容量を達成することが Barron らによって証明されている．ここに，復号問題はモデル選択と本質的に同じ問題となり原理と関連する．この符号について，辞書を Bernoulli 分布から生成した場合を考察し，計算量を無視して最適な復号法を用いた場合に通信路容量を達成することを証明した (Takeishi, Kawakita, and Takeuchi 2014). さらに，圧縮センシングとの関連について考察を進め，Bayes 最適 Approximate Message Passing (AMP)[A3]を用いると，Barron らによる従来の効率的復号法[A8]よりも高い伝送速度のもとで復号が可能であることを実験的に示した．

［参考文献］

[A1] Amari and Nagaoka, *Methods of Information Geometry*, AMS & Oxford University Press, 2000.

[A2] Barron and Cover, “Minimum complexity density estimation,” *IEEE Trans. IT*, vol. 37, pp. 1034-1054, 1991.

[A3] Donoho, Javanmard, and Montanari, ``Information-Theoretically Optimal Compressed Sensing via Spatial Coupling and Approximate Message Passing," *IEEE Trans. Inf. Theory*. vol. 59, no. 11, pp. 7434-7464, Nov., 2013.

[A4] Grunwald et al. (Editors), *Advances in Minimum Description Length*, MIT Press, 2005

[A5] Grunwald, *The Minimum Description Length Principle*, MIT Press, 2007.

[A6] Jacquet and Szpankowski, “Markov types and minimax redundancy for Markov sources,” *IEEE trans. IT*, Vol. 50, No. 7, July 2004.

[A7] Joseph and Barron, “Least Squares Superposition Codes of Moderate Dictionary Size Are Reliable at Rates up to Capacity,” *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2541-2557, May 2012.

[A8] Joseph and Barron, “Fast sparse superposition codes have near exponential error probability for R < C,” *IEEE Trans. IT*, vol. 60, no. 2, pp. 919–942, Feb. 2014.

[A9] Kontkanen and Myllymaki, “A linear-time algorithm for computing the multinomial stochastic complexity,” *Information Processing Letters,* 103 pp. 227-233, 2007.

[A10] Rissanen, “Modeling by shortest data description,” *Automatica*, vol. 14, pp. 465-471, 1978.

[A11] Rissanen, “Fisher information and stochastic complexity,” *IEEE trans. IT*, vol. 40, pp. 40-47, 1996.

[A12] Takeuchi and Barron, “Asymptotically minimax regret by Bayes mixtures,” *Proc. 1998 IEEE Int. Symp. IT*, p. 318, 1998.

[A13] Takeuchi and Kawabata, “Exponential curvature of Markov models,” *Proc. 2007 IEEE International Symposium on Information Theory*, 2007.

[A14] Weinberger, Rissanen and Feder, “A universal finite memory source,” *IEEE trans. Inform. Theory*, Vol. 41. No. 3, pp. 643-652, 1995.

[A15] Xie and Barron, “Asymptotic minimax regret for data compression, gambling and prediction,” *IEEE trans. Inform. Theory*, vol. 46, no. 2, pp. 431-445, 2000.

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　4　件）

① Y. Takeishi, M. Kawakita, J. Takeuchi, "Least Squares Superposition Codes with Bernoulli Dictionary are Still Reliable at Rates up to Capacity," *IEEE Transactions on Information Theory*, Vol. 60, No. 5, pp. 2737-2750, May 2014.

② J. Takeuchi, T. Kawabata, A. R. Barron, "Properties of Jeffreys Mixture for Markov Sources," *IEEE Transactions on Information Theory*, Vol. 59, No. 1, pp. 438-457, January 2013.

〔学会発表〕（計　22　件）

① J. Takeuchi, A. R. Barron, "Stochastic complexity for tree models," *Proc. of 2014 IEEE Information Theory Workshop*, pp. 223-227, Hobart, Tasmania, Australia, November 2-5, 2014

② J. Takeuchi, A. R. Barron, "Asymptotically minimax regret for models with hidden variables," *Proc. of 2014 IEEE International Symposium on Information Theory*, pp. 3037-3041, Honolulu, HI, USA, June 29 - July 4, 2014

③ J. Takeuchi, A. R. Barron, "Asymptotically Minimax Regret by Bayes Mixtures for Non-exponential Families," *Proc. of 2013 IEEE Information Theory Workshop*, pp.

204-208, Seville, Spain, September 9-13, 2013.

④ Y. Takeishi, M. Kawakita, J. Takeuchi, "Least Squares Superposition Codes with Bernoulli Dictionary are Still Reliable at Rates up to Capacity," *Proc. of 2013 IEEE International Symposium on Information Theory*, pp. 1396-1400, Istanbul, Turkey, July 7-12, 2013.

⑤ S. Yamauchi, M. Kawakita, J. Takeuchi, "Botnet Detection based on Non-negative Matrix Factorization and the MDL Principle," *Proc. of the 19th International Conference on Neural Information Processing*, Doha, Qatar, Nov. 12-15, 2012

⑥ M. Tsurusaki, J. Takeuchi, "Constant Markov Portfolio and Its Application to Universal Portfolio with Side Information," *Proc. of 2012 IEEE International Symposium on Information Theory*, Boston, USA, July 1-6, 2012.

〔図書〕（計　1　件）

① J. Takeuchi, "An introduction to the minimum description length principle," in *A Mathematical Approach to Research Problems of Science and Technology*, pp. 279-296, Springer, 2014. (book chapter)

６．研究組織

(1)研究代表者

竹内　純一　（TAKEUCHI JUNICHI）
九州大学・システム情報科学研究院・教授
研究者番号：80432871

(3)連携研究者

小野　廣隆　（ONO HIROTAKA）
九州大学・経済学研究院・准教授
研究者番号：00346826

(4)研究協力者

Andrew R. Barron
Yale University・Dept. of Statistics・教授