

非定常情報源に対する resetting 分布を用いたモデル系列推定

櫻井瑛一*

Eiichi Sakurai

山西 健司†

Kenji Yamanishi

Abstract: データ生成源のモデルが時間とともに変化して発生したときに、その非定常なデータからそのモデルの系列をいかに推定するかという問題を考える。本発表ではモデル系列の構造として switching 型と resetting 型があることを紹介する。特に、resetting 分布を用いて、モデル遷移に新しい構造を入れることにより、区間定常的なモデルの遷移を効率的に抽出できることを示す。しかも、その推定に関する情報論的限界が switching 型のそれよりも小さくなることを示す。

Keywords: 統計的モデル選択, Dynamic Model Selection

1 発表の概略

1.1 背景

本発表ではデータ生成源の確率モデルが時間とともに変化しているような非定常情報源を考え、その変化を推定する問題を考える。そして、各確率モデルをたとえば、モデルパラメータの次数などによるモデル複雑度の指数で表現することにする。このとき、データ生成源の変化はモデルの系列として表現でき、それを推定すればよい。本発表ではこの問題を [6, 7] にならい動的モデル選択 (*Dynamic Model Selection*, DMS と略す) と呼ぶ。非定常性を示すデータは、サービスの場面ではたとえば顧客の商品の購買行動や来店行動がある。このとき非定常性を生じる要因は、マクロには景気の変動や気温の変化、商品の人気などがあり、ミクロには日々の温度、商品の配置換えなど様々に存在する。したがって、このような非定常なデータから変化の時点とその生成構造の変化の仕方を推測することは、変化要因を推定し、サービスの質の向上や需要を予測しサービスを最適化する上で重要な問題である。

1.2 従来手法

非定常性を示すデータに対して、その性質に対応する手法として大きく分け二つのタイプがある。ひとつは、

よい構造が切り替わっていくと考えるものである。これには AIC-BIC ジレンマを解消したデータの数とともに最適モデル次数を変える Tim van Erven et al. による switching の考え方 [1] や最適な expert が切り替わるときに、予測の重みをうまく学習させることで対応する tracking the best expert [2], モデルが切り替わるとしそのモデルの系列を動的計画法で選択する動的モデル選択 [6, 7] がある。

他方には区間定常的な情報源から出現したデータに対して、データ圧縮を行う研究がある。これには [5] や [4] がある。これらで重要な点は、構造変化が生じた時、つまり定常な区間が終了し次の定常な区間に移行した時点から推定を行いなおすことを考え、すべての変化の仕方に重み付けすることで、隠れた区間定常性に対処するところにある。

1.3 本発表の手法概略

本発表では [6, 7] にて提案された記述長最小原理 (*Minimum Description Length principle*, MDL 原理と略す) [3] に基づきモデル系列を選択する DMS 規準を用いる。MDL 原理が (モデル自身の記述長) と (モデルの下でのデータの記述長) の和を最小にするモデルを選択することを考えていたが、この規準では、モデルをモデルの系列と読み替える、すなわち、(モデル系列自身の記述長) と (モデル系列の下でのデータの記述長) の和を最小にするモデル系列を選択することを考える。本発表では、この (モデル系列の下でのデータの記述長) の部分をどのように定めれば、記述長を小さく、そして、計算効率性を落とさず定義できるかを考える。

そのために、本発表では resetting 分布と switching

*産業技術総合研究所サービス工学研究センター
135-0064 東京都江東区青海 2-3-26
Center for Service Research, National Institute of Advanced Industrial Science and Technology (AIST)
2-3-26 Aomi, Koto-ku, Tokyo 135-0064, Japan

†東京大学
113-8656 東京都文京区本郷 7-3-1
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

分布というモデル系列の下でのデータ記述法を二種類紹介する。前者は変化点ごとに過去のデータを忘れ、推定を構成しなおす分布であり、後者は過去のデータから構成された予測分布がよくなる時点でモデルが切り替わる分布である。両者のアイデアは先の既存の手法にあり、resetting 分布のアイデアは前節の手法の後者、すなわち区間定常的な情報源に対するデータ圧縮手法にある。他方 switching 分布は前節の手法の前者、すなわち、よい構造が現れた時点で構造が切り替わっていくという考え方にある。このとき、両者の間の大きな差は、モデルのパラメータ推定にどれだけ過去に依存しているかにある。

本発表では resetting 分布と switching 分布を用いた DMS を設計し解析する統一フレームワークを与える。具体的には以下の二つを発表する。

- resetting 分布と switching 分布を DMS に適用し、計算効率性を落とさない様な DMS 規準最小化アルゴリズムの要点を紹介する
- 情報論的限界を比較可能な形で導出し、理論的実験的比較を行う

参考文献

- [1] S. de Rooij and T. van Erven. Learning the switching rate by discretising bernoulli sources online. In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2009.
- [2] M. Herbster and M. K. Warmuth. Tracking the best expert. In *Proceedings of the 12th International Conference on Machine Learning*, pp. 286–294, 1995.
- [3] J. Rissanen. *Information and Complexity in Statistical Modeling*. Springer, 2010.
- [4] G. I. Shamir and N. Merhav. Low complexity sequential lossless coding for piecewise stationary memoryless sources. *IEEE Transactions on Information Theory*, Vol. 45, pp. 1498–1519, 1999.
- [5] F. M. J. Willems. Coding for a binary independent piecewiseidentically-distributed source. *IEEE Transactions on Information Theory*, Vol. 42, pp. 2210–2217, 1996.
- [6] K. Yamanishi and Y. Maruyama. Dynamic syslog mining for network failure monitoring. In

Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 499–508, 2005.

- [7] K. Yamanishi and Y. Maruyama. Dynamic model selection with its applications to novelty detection. *IEEE Transactions on Information Theory*, Vol. 53, pp. 2180–2189, 2007.