

## 時系列に対するグラフィカルモデルの動的潜在構造最適化

数理情報学専攻 48216217 佐藤 瞭

指導教員 山西 健司 教授

## 1 はじめに

グラフィカルモデルは確率モデルをグラフの形で記述したものであり、確率モデルを簡潔に可視化できる、グラフ構造から変数間の条件付独立性等のモデルの性質が分かるという利点がある。

値が観測されない確率変数である潜在変数を含むグラフィカルモデルに関する先行研究や状態空間モデルの学習に関する先行研究は多数存在する一方で、従来の研究において、潜在変数の次元が変化する場合の手法は考えられていない。しかしながら、潜在変数の次元の変化検知により、一時的な施策の効果が及んだ範囲、期間や、社会情勢の変化等による地域間関係の変化等を分析できる可能性がある。本研究では、 $d_x$  次元の観測変数の時系列  $x^{1:T} = x^1, \dots, x^T$  と  $d_z$  次元の潜在変数の時系列  $z^{1:T} = z^1, \dots, z^T$  に対し、潜在変数を含む次式で表される多変量自己回帰モデルを考える。

$$\begin{aligned} z^t &= Az^{t-1} + w, \\ x^t &= Bx^{t-1} + Cz^{t-1} + v, \\ z^1 &= \mu^0 + u, \\ w &\sim \mathcal{N}(w|\mathbf{0}, \Gamma), v \sim \mathcal{N}(v|\mathbf{0}, \Sigma), u \sim \mathcal{N}(u|\mathbf{0}, P^0). \end{aligned}$$

本研究では、このモデルを用いて、グラフィカルモデルとしての表現力を確保しながら潜在構造を動的に推定する手法を記述長最小原理 (MDL 原理) の立場から新規に提案する。

## 2 準備:MDL 原理の立場によるモデル選択

MDL 原理の立場におけるモデル選択規準として正規化最尤符号長 (NML 符号長) が挙げられる。しかし、潜在変数モデルに対して直接 NML 符号長を計算するのは困難なので次の計算方法を用いる。

## 2.1 分解型正規化最尤符号長 (DNML 符号長)

確率分布が  $p(X, Z; \theta) = p(X|Z; \theta_X)p(Z; \theta_Z)$  と分解できるモデル  $M$  に対する DNML 符号長 [5] は次式で定義される。ここで  $L_{\text{NML}}$  は NML 符号長を示す。

$$\begin{aligned} L_{\text{DNML}}(x^{1:T}, z^{1:T}, M) \\ = L_{\text{NML}}(x^{1:T}|z^{1:T}, M) + L_{\text{NML}}(z^{1:T}, M). \end{aligned}$$

## 3 モデルの学習

潜在変数の次元  $d_z$  の推定値の候補毎に、観測変数のみを考える初期モデルの学習と、初期モデルの情報を利用したモデル全体の学習の 2 段階からなる学習を行う。

## 3.1 初期モデルの学習

まずは観測変数のみを含む、次のモデルを考える。

$$x^t = Bx^{t-1} + v_x, \quad v_x \sim \mathcal{N}(v_x|\mathbf{0}, \Sigma_x). \quad (1)$$

グラフィカルモデルとして記述した際に変数同士の影響の有無が鮮明に表現されるようにするために、パラメータ  $B$  の学習には正則化をかける。また、ノイズは各変数で独立であるという仮定し、式 (1) を各変数毎の線形回帰の問題に分解して考える。本研究では Luckiness NML 符号長 [1] を利用してパラメータを推定する。

## 3.2 モデル全体の学習

3.1 節で得られたパラメータの推定値を初期値として、EM アルゴリズムによってモデル全体のパラメータを推定する。E-step における  $p(z^t|x^{1:T})$  の計算は、動的計画法の要領で効率よく行うことができる。M-step におけるパラメータの推定では、 $B$  の推定値をスパースに保つために、初期モデルで値がゼロと推定された要素は EM アルゴリズムによる更新後も値がゼロのままであるという制約を導入する。

## 4 モデルの動的潜在構造最適化

各時刻において  $d_z$  の推定値の候補を考え、それぞれに対してモデルを学習し、 $d_z$  の推定規準及び変化検知の規準を計算することで  $d_z$  の推定時系列を構成する。

## 4.1 潜在変数の次元の推定規準

潜在変数の次元  $d_z$  の推定には DNML 符号長を用いる。本研究では、モデル全体の DNML 符号長を各変数毎の符号長の和で近似する、因子化 NML 符号長 [3] と呼ばれるテクニックを用いた。また、各変数毎の符号長には線形回帰モデルの符号長 [2] を用いた。

## 4.2 潜在変数の次元の変化検知

時刻  $t$  での潜在変数の次元  $d_z$  の変化の有無は次式で定義される MDL 変化統計量 [4] によって判定する。

$$\begin{aligned} \Phi_t(x^{1:n}) &= \mathcal{L}(x^{t-h:t+h}) \\ &\quad - \{\mathcal{L}(x^{t-h:t}) + \mathcal{L}(x^{t+1:t+h}) + 2h\epsilon\}. \end{aligned}$$

$\mathcal{L}$  はデータ及びモデルの記述長であり,  $\varepsilon, h > 0$  はハイパーパラメータである. MDL 変化統計量は, データ  $x^{t-h:t+h}$  を 1 つの確率モデルを用いて表現する場合と比べ, 時刻  $t$  の前後でデータを区切り別々のモデルで表現するとどれ位記述長を短くできるか測るものである.

## 5 数値実験

### 5.1 人工データ: $d_z$ の推定精度

まず,  $d_z$  の真の値が変化しない時系列を用いて  $d_z$  の推定精度を確認した. モデル選択規準として符号長による規準 (MDL) 及び AIC, BIC を比較した. 結果を図 1 に示す. MDL 規準は AIC, BIC より高い精度を示す一方, データ長  $T$  が大きい場合は精度が低下した.

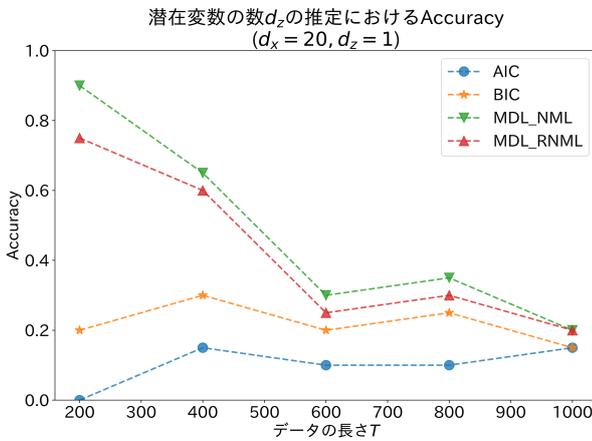


図 1. 潜在変数の次元の推定精度.

### 5.2 人工データ: $d_z$ の変化検知精度

次に,  $d_z$  の値が時刻と共に変化する時系列を用いて  $d_z$  の変化検知の精度を確認した. 詳細な結果は紙面の都合上省略する.  $d_z$  の値が  $0 \rightarrow 1$  に増加する変化の検知には MDL 変化統計量が高い精度を示す一方,  $d_z$  の値が  $1 \rightarrow 0$  に減少する変化は MDL 変化統計量では検知できず, MDL 変化統計量の符号長の部分に BIC を代入したものが不安定ながらも比較的高精度を示した.

### 5.3 実データ

2012 年から 2013 年における 23 か国の株式インデックスの時系列 [6] に対し, 各時刻で  $d_z$  の値を推定した.  $d_z$  の増加判定には MDL 変化統計量を, 減少判定には MDL 変化統計量の符号長を BIC で置き換えたものを用いた結果, 全ての期間において  $d_z = 1$  であるという推定が得られた. 学習したモデルのパラメータを図 2 に示す. 初期モデルの学習での正則化によりパラメータ  $B$  の推定値はスパースになっており, グラフィカルモデルとしての表現力を確保できている. また, パラメータ  $C$  の推定値から, 潜在変数はヨーロッパ全体における景気動向を示す指標であるという解釈ができる.

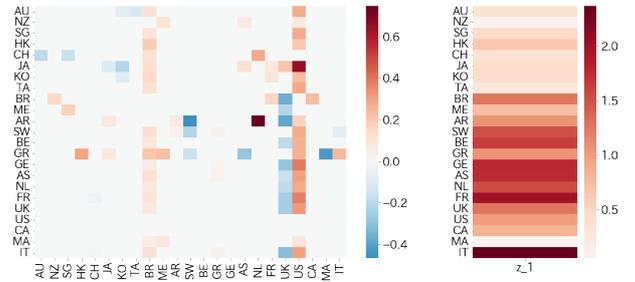


図 2. パラメータ  $B$  (左),  $C$  (右) の推定結果.

## 6 結論

本研究では, 潜在変数を含む多変量自己回帰モデルに対し, グラフィカルモデルとしての表現力を確保しながら潜在変数の次元推定及び次元変化の検知を行う手法を MDL 原理の立場から新規に提案した.

数値実験において提案手法は AIC や BIC と比べて概ね高い精度を示したが, いくつかの場合では低い精度を示した. この原因として, 潜在変数値の推定に関する記述長を考慮していないことや, 潜在変数値の推定精度が低い可能性, パラメータ  $B$  の構造を初期モデルの学習の時点で固定していることが考えられる. 改善策として, 潜在変数値の推定に関する記述長の導入や, EM アルゴリズムへの正則化の導入が挙げられる. 正則化に関しては, 4.1 節で考えた符号長の代わりに Luckiness NML 符号長の新たな導出が必要となる.

また, 実データを用いた分析により, 提案手法がグラフィカルモデルとしての表現力を保った学習方法となっていることを確かめ, 潜在変数を考えることによる知識発見の一例を示した. より広分野, 長期間にわたる実データへの提案手法の応用も今後の課題である.

## 参考文献

- [1] K. Miyaguchi and K. Yamanishi. High-dimensional penalty selection via minimum description length principle. *Machine Learning*, 107:1283–1302, 2018.
- [2] J. Rissanen. MDL denoising. *IEEE Transactions on Information Theory*, 46(7):2537–2543, 2000.
- [3] T. Silander, T. Roos, P. Kontkanen, and P. Myllymäki. Factorized normalized maximum likelihood criterion for learning Bayesian network structures. In *Proceedings of the 4th European workshop on probabilistic graphical models*, pp. 257–272, 2008.
- [4] K. Yamanishi and S. Fukushima. Model change detection with the MDL principle. *IEEE Transactions on Information Theory*, 64(9):6115–6126, 2018.
- [5] K. Yamanishi, T. Wu, S. Sugawara, and M. Okada. The decomposed normalized maximum likelihood code-length criterion for selecting hierarchical latent variable models. *Data Mining and Knowledge Discovery*, 33(4):1017–1058, 2019.
- [6] M. Zorzi and R. Sepulchre. AR identification of latent-variable graphical models. *IEEE Transactions on Automatic Control*, 61(9):2327–2340, 2016.