

非定常データからのネットワーク構造変化検出

早矢仕 裕*
Yu Hayashi

山西 健司†
Kenji Yamanishi

Abstract: 非定常な時系列データから、変数間の依存関係（ネットワーク構造）の変化を検出する問題を扱う。本稿では、データの従う確率モデルにグラフィカルモデルを導入して、非定常データからグラフィカルモデルの系列を学習することでネットワーク構造の変化を検出する。Xuan and Murphy の手法や Robinson and Hartemink の手法など、このような方針のもとで従来に提案された手法の概説と、動的モデル選択に基づく手法の提案を行う。さらに手法のマーケティングにおける広告効果測定への応用について述べる。

Keywords: ネットワーク構造変化検出, グラフィカルモデル, 動的モデル選択

1 はじめに

本稿では、非定常な時系列データからのネットワーク構造変化検出について扱う。ネットワーク構造変化検出とは、データの変数同士が互いに関係を持つ時系列データが与えられ、さらに時間発展するにつれて新たに関係が生起・消滅するような場合に、各時刻での変数間の関係とその変化を明らかにする問題である。このような構造を有するデータの例として、マーケティングデータ、経済時系列データ、センサーデータなどが挙げられる。

はじめに、本稿での問題設定について述べる。まず、各時刻でのデータがグラフ構造 G を持った確率モデル（グラフィカルモデル） $P(x : G)$ から生成され、さらにグラフィカルモデルのグラフ構造が時間と共に $G_1 \rightarrow G_2 \rightarrow \dots \rightarrow G_\ell$ と変化していく状況を考える。このとき、ネットワーク構造変化検出は、データ列 $x^T = x_1 x_2 \dots x_T$ からグラフィカルモデルの系列 $(G_1, G_2, \dots, G_\ell)$ と変化点 $(t_1, t_2, \dots, t_{\ell-1})$ の組を推定する問題として定義される。

2 関連研究の紹介

本章では、ネットワーク構造変化検出に関する既存研究の概説を行う。

2.1 MCMC を用いた手法

ネットワーク構造変化を検出する手法の一つとして、マルコフ連鎖モンテカルロ法 (MCMC) を利用した手法がある。データの確率モデルとしてグラフィカル・ガウシアン・モデルを用いた手法が Talih and Hengartner [4] によって、ベイジアンネットワークを用いた手法が Robinson and Hartemink [3] によって提案されている。

これらの手法では、データ列 x^T が与えられたもとでモデル系列 $G = (G_1, G_2, \dots, G_\ell)$ とその変化点 $t = (t_1, t_2, \dots, t_{\ell-1})$ に対する事後分布 $P(G, t | x^T)$ を計算することでこれらの推定を行う。このとき G, t のとりうる候補の数は非常に大きいため、MCMC の一手法であるメトロポリス・ヘイスティングス法により事後分布からのサンプリングを行っている。

2.2 動的計画法を用いた手法

動的計画法によりモデル系列の推定を行う手法として、Xuan and Murphy [5] による手法が提案されている。この手法において、変化点検出には Fearnhead and Liu [1] の手法が利用されている。Fearnhead and Liu の手法は、モデルのクラス (G_1, \dots, G_n) とモデルに関する事前分布 $p(G)$ が与えられたときに、データ列 x^T の周辺尤度を最大化する変化点の組を動的計画法により探索する。

Xuan and Murphy の手法は、初めにヒューリスティクスによりモデルクラスを構成し

- モデルクラスと一様なモデルの事前分布が与えられたもとで Fearnhead and Liu の手法によりデータからの変化点検出を行う

*東京大学大学院 情報理工学系研究科, 113-8656 東京都文京区本郷 7-3-1, yu_hayashi@mist.i.u-tokyo.ac.jp, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8656 Japan

†東京大学大学院 情報理工学系研究科, 113-8656 東京都文京区本郷 7-3-1, yamanishi@mist.i.u-tokyo.ac.jp, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8656 Japan

- 変化点によって分割されたデータの各区間について、モデルを学習することでモデルクラスを更新する

という2つの操作を収束するまで繰り返すことにより、モデル系列と変化点の推定を行う。

3 現在の取り組み

3.1 動的モデル選択に基づく手法

現在までの取り組みとして、ネットワーク構造変化検出の問題に対して動的モデル選択 [6] の理論に基づく手法を提案している。動的モデル選択とは非定常なデータからモデルの構造変化を追跡するための理論であり、MDL 原理 [2] の枠組みの中で解かれてきた。

提案手法では各時刻においてモデルがマルコフ過程 $P(G_t|G_{t-1})$ に従って変化する状況を仮定する。このとき与えられたデータ列 x^T に対し、以下の規準を最小化するモデル系列を出力する。

$$\ell(x^T : G^T) \stackrel{\text{def}}{=} \sum_{t=1}^T -\log P(x_t | \hat{\theta}_{G_t}^{(t-1)}) + \sum_{t=1}^T -\log P(G_t | G_{t-1}).$$

ただし、 $\hat{\theta}_{G_t}^{(t-1)}$ は $x^{t-1} = x_1 \dots x_{t-1}$ から推定したモデル G_t のパラメータである。この規準の第一項はモデル系列が与えられた下でのデータの符号長、第二項はモデル系列自体の符号長に対応している。すなわちこれはデータ圧縮の意味で最適なモデル系列を見つけることに他ならない。この規準は t に対して逐次計算できるため動的計画法が適用できる。

3.2 マーケティングへの応用

また、ネットワーク構造変化検出の応用として、マーケティングにおける広告効果測定に取り組んでいる。ある製品に関するデータとして、広告出稿量、ブログ投稿数、Web 検索数、販売台数などの変数を持つデータが与えられたとする。このとき提案手法により、各変数がどのような関係を持つか、またそれらの関係がいつどのように変わるかを明らかにすることで、各期間で出稿された広告が市場に与える影響を明らかにすることを目的としている。現在までに実データから変数間の関係変化が捉えられることを確認し、その有効性について検証している。

4 おわりに

本稿では、ネットワーク構造変化検出における既存手法の概説と、動的モデル選択に基づく手法の提案を行っ

た。また、ネットワーク構造変化検出の応用例として、マーケティングにおける広告効果測定への応用について示した。

参考文献

- [1] P. Fearnhead and Z. Liu. Online inference for multiple changepoint problems. *Journal Of The Royal Statistical Society Series B*, 69(4):589–605, 2007.
- [2] J. Rissanen. *Information and complexity in statistical modeling*. Springer-Verlag, 2007.
- [3] J. W. Robinson and A. J. Hartemink. Learning non-stationary dynamic bayesian networks. *Journal of Machine Learning Research*, 11:3647–3680, 2010.
- [4] M. Talih and N. Hengartner. Structural learning with time-varying components: tracking the cross-section of financial time series. *Journal Of The Royal Statistical Society Series B*, 67(3):321–341, 2005.
- [5] X. Xuan and K. Murphy. Modeling changing dependency structure in multivariate time series. In *Proceedings of the 24th International Conference on Machine Learning*, pages 1055–1062, 2007.
- [6] K. Yamanishi and Y. Maruyama. Dynamic model selection with its applications to novelty detection. *IEEE Transactions on Information Theory*, 53(6):2180–2189, 2007.