

グラフフーリエ変換を用いた グラフ上のパターンに関する数理的解析

数理情報学専攻 48-176226 森下 真幸

指導教員 合原 一幸 教授

1 はじめに

グラフ信号がなす空間パターンは各頂点間の相関などを示唆する重要な研究対象である。本研究では、グラフ信号による空間パターンの非線形性の検定を目的に、非線形時系列解析のサロゲートデータ法のグラフ信号への拡張手法を提案した。またグラフ信号に適用可能な非線形性指標を提案し、数値実験により性質を確認した。

2 基礎理論

2.1 サロゲートデータ法

サロゲートデータ法 [1] は時系列データに対する統計的仮説検定の枠組みである。手順は次の通りである。

- 手順 1. 帰無仮説 H_0 と統計量, 有意水準 α を決定する。
- 手順 2. 検定したい時系列データを元に, H_0 を満たすようなデータをランダムに S サンプル生成する。生成されたデータはサロゲートデータと呼ばれる。
- 手順 3. サロゲートデータの検定統計量を算出し, H_0 下での検定統計量の分布を推定する。
- 手順 4. 元のデータに対しても検定統計量を算出し, 手順 3. の分布と比較し, H_0 を棄却あるいは採択する。

なお, 具体的なサロゲートデータの生成手法は H_0 に依存して決まる。また両側検定の場合には $\alpha \geq \frac{2}{S+1}$, 片側検定の場合には $\alpha \geq \frac{1}{S+1}$ を満たす必要がある。

2.2 グラフ信号処理

グラフ信号処理 [2] は, ネットワーク構造を積極的に用いて既存の信号処理を拡張する試みである。

2.2.1 グラフラプラシアンとグラフフーリエ変換

頂点集合 \mathcal{N} , 辺集合 $\mathcal{E} \subset \mathcal{N} \times \mathcal{N}$, 重み関数 $\mathcal{W} : \mathcal{N} \times \mathcal{N} \rightarrow \mathbb{R}_{\geq 0}$ の有限の重み付き無向連結グラフ $\mathcal{G} = \{\mathcal{N}, \mathcal{E}, \mathcal{W}\}$ を考える。頂点数 $N = |\mathcal{N}|$ とする。また, $(n_i, n_j) \in (\mathcal{N} \times \mathcal{N}) \setminus \mathcal{E}$ の時, $\mathcal{W}(n_i, n_j) = 0$ とする。

\mathcal{G} の隣接行列を $W = (W(n_i, n_j))_{ij}$ ($n_i, n_j \in \mathcal{N}$), 次数行列を $D = \text{diag} \left(\left(\sum_j W(n_i, n_j) \right)_i \right)$ とする。また, グラフラプラシアン L を $L := D - W$ で定義する。

定義 1. (グラフフーリエ変換 (Graph Fourier Transform, GFT)) グラフ \mathcal{G} の頂点 \mathcal{N} 上のグラフ信号

$x = (x_1, x_2, \dots, x_N)^T \in \mathbb{R}^N$ に対し, GFT とその逆変換 Inverse GFT (GFT^{-1}) を以下の通り定義する:

$$\text{GFT}[x] := V^T x, \quad \text{GFT}^{-1}[\tilde{x}] := V \tilde{x}.$$

なお, V は L の対角化行列であり, $\lambda \in \mathbb{R}^N$ を L の固有値からなるベクトルとして $L = V \text{diag}(\lambda) V^T$ を満たす。

2.2.2 グラフ上の定常性と Wiener-Khinchin の定理

グラフ信号についても, その定常性としてグラフ広義定常性 (Graph Wide Sense Stationarity, GWSS) を考えることができる。ここではその定義 [3] については省略し, 同値で有用な定理を紹介する。

定理 1. (グラフ上の Wiener-Khinchin の定理 [3]) グラフ \mathcal{G} 上の確率変数 $\{X_n\}$ ($n \in \mathcal{N}$) が GWSS であることと, 以下の等式が成立することは同値である:

$$\Sigma = V \text{diag}(p) V^T \quad (\Sigma_{nm} = \text{Cov}[X_n, X_m], p \in \mathbb{R}^N),$$

よって, グラフ信号が GWSS であれば, 以下のよう
にパワースペクトル密度 (PSD) を表すことができる:

$$p = \mathbb{E} \left[\left| V^T (X - \mathbb{E}[X]) \right|^2 \right] \quad (X = (X_n)_{n \in \mathcal{N}}).$$

3 提案手法

3.1 サロゲートデータの生成

グラフ信号の非線形性の検定を目的として,

H_0 : 信号はグラフ上の線形ガウス過程から生成された。

を考える。このとき, 2 次までのモーメントでグラフ上の過程 $\{X_n\}_{n=1}^N$ を記述できる。また, グラフ上の Wiener-Khinchin の定理から, 分散共分散行列とグラフ上の PSD が等価なので, PSD 不変なランダムな変換として, $U = V^T A V$ (A : ランダム直交行列) による,

$$y = \text{GFT}^{-1} [U \text{GFT}[x - \bar{x}]] \quad (\bar{x} = \sum_{n=1}^N x_n / N)$$

を考え, サロゲートデータ $y \in \mathbb{R}^N$ を得る手法を GFT サロゲートデータ法と呼ぶこととする。

ところで, 本手法は平均値を保存しない。そこで, 平均値も保存する変換として, Amplitude Adjusted GFT (AAGFT) サロゲートデータ法も提案した。

3.2 評価指標

提案手法によるサロゲートデータ生成の後, 検定を行うには統計量が必要である. そこで, 以下の Volterra 多項式による指標を提案する.

グラフ信号に対する次数 D , メモリ M の $y \in \mathbb{R}^N$ 上の Volterra 多項式 p_D^M を以下のように定義する:

$$\begin{aligned} p_D^M(y) := & a_0 + a_1(I-L)^\dagger y + a_2(I-L)^\dagger{}^2 y + \dots \\ & + a_M(I-L)^\dagger{}^M y + a_{M+1} \odot^2 ((I-L)^\dagger y) \\ & + a_{M+2}((I-L)^\dagger y) \odot ((I-L)^\dagger{}^2 y) \\ & + \dots + a_P \odot^D ((I-L)^\dagger{}^M y). \end{aligned}$$

ただし, \dagger は疑似逆行列を, \odot はアダマール積, つまり要素ごとの積を表し, 特に $\odot^D y = (y_n^D)_{n=1}^N$ である. また, $P = \frac{(M+D)!}{M!D!}$ で Volterra 多項式の項数を表す.

グラフ信号 x について $x = p_D^M(x)$ となるように $\{a_p\}_{p=0}^P$ を最小 2 乗法で推定し, その結果を \hat{p}_D^M とすると, グラフ信号の非線形性の指標 ρ_D^M は $\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n$ を用いて以下のように定義できる:

$$\begin{aligned} \rho_D^M := & K[p_D^M] - K[p_1^M] \quad (D \geq 2), \\ \left(K[p_D^M] := & 1 - \frac{\|x - \hat{p}_D^M(x)\|^2}{\sum_{n=1}^N (x_n - \bar{x})^2} \quad (D \geq 1) \right). \end{aligned}$$

3.3 既存手法

GFT に基づいたサロゲートデータの生成手法として, 既に Sign Randomization[4] という手法が提案されている. この手法では ± 1 をランダムに取る対角行列 U を用いて, サロゲートデータを生成する. 以下では, この手法を SR サロゲートデータ法と表すことにする.

4 数値実験と考察

ランダムグラフ上でダイナミクスに基づくグラフ信号を用い, 数値実験により提案事項の性質を評価した.

ランダムグラフは, Erdős-Rényi, Barabási-Albert (BA) の 2 モデルについて頂点数 100 で実験を行った.

ダイナミクスのモデルには, (a) 拡散方程式 $\frac{dX(t)}{dt} = D_X LX(t)$, (b) Allen-Cahn 方程式 $\frac{dX(t)}{dt} = D_X LX(t) + u(1-u)(u - \frac{1}{2})$, また Turing パターンを持つことで知られる以下の (c) 三村-Murray (MM) モデルを用いた:

$$\begin{cases} \frac{dX(t)}{dt} = D_X LX(t) + \left(\frac{1}{9}(35 + 16X - X^2) - Y\right) X, \\ \frac{dY(t)}{dt} = D_Y LY(t) - \left((1 + \frac{2}{5}Y) - X\right) Y. \end{cases}$$

ただし, 拡散定数 D_X は 0.06 で固定, D_Y は 0.6 と 1.8 の 2 通りを試した. いずれも $X(t)$ のみを利用した.

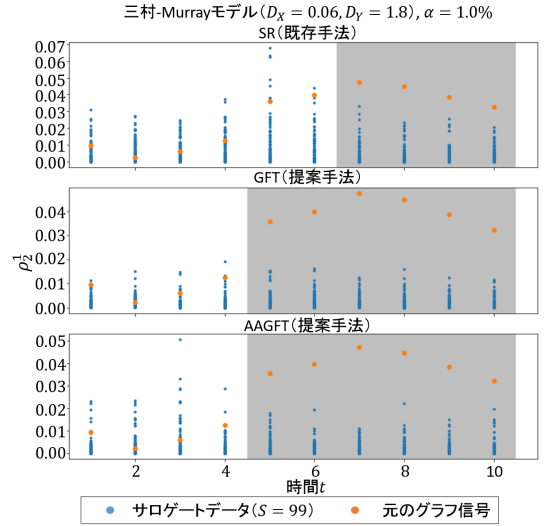


図 1. BA グラフ上の MM モデル ($D_Y = 1.8$) によるグラフ信号の検定結果の時間推移. 背景がグレーの時点で H_0 が棄却されている.

実験の結果, ρ_D^M は概ね, $(c : D_Y=1.8) > (a) \approx (c : D_Y=0.6) > (b)$ となり, 空間パターンの非一様性を定量化している様子であった. そこで, ρ_2^1 で既存手法, 提案手法による 1% 水準の検定を実施し, 以下が分かった.

- 図 1 の結果より, 空間パターンの非線形はいずれの手法でもある程度検定できていると考えられる.
- ただし, Turing パターンに対して既存手法よりも提案手法の方が H_0 を棄却できていると考えられる.

5 おわりに

本研究では, グラフ信号がなす空間パターンの非線形性の検定を目的とし, その検定手法と用いる指標の提案を行った. 数値実験により, 提案指標がパターンの非一様性を定量化している点, 提案手法が既存手法より優れていると考えられることが明らかになった.

参考文献

- [1] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, “Testing for nonlinearity in time series: the method of surrogate data,” *Physica D: Nonlinear Phenomena*, vol. 58, no. 1, pp. 77–94, 1992.
- [2] A. Sandryhaila and J. M. F. Moura, “Discrete signal processing on graphs,” *IEEE Transactions on Signal Processing*, vol. 61, pp. 1644–1656, April 2013.
- [3] N. Perraudin and P. Vandergheynst, “Stationary signal processing on graphs,” *IEEE Transactions on Signal Processing*, vol. 65, pp. 3462–3477, July 2017.
- [4] E. Pirondini, A. Vybornova, M. Coscia, and D. V. D. Ville, “A spectral method for generating surrogate graph signals,” *IEEE Signal Processing Letters*, vol. 23, pp. 1275–1278, Sep. 2016.