

系統樹ガウス過程因子モデルとディリクレ過程によるクラスタリングを用いた結合型種分布モデル

理化学研究所 AIP センター 経済経営情報融合分析チーム 特別研究員 城田慎一郎
国立環境研究所 生物・生態系環境研究センター 特別研究員 深谷肇一 吉川徹朗

近年、群集生態学において、多様な種の分布パターンを同時にモデル化する、結合型種分布モデル (Joint Species Distribution Model, JSMD) が注目を集めている。数百種類の種の分布パターンを同時にモデリングするアプローチも提案されはじめており、Taylor-Rodríguez et al. (2016) は、共変量で捉えられない種間の類似構造にディリクレ過程によりクラスタリング構造を導入した因子モデルを提案した。また、Shirota et al. (2019) では、上記モデルを拡張し、観測された位置情報に基づいた、空間ガウス過程を導入している。本研究では、位置情報の代わりに、観測された植物の系統樹情報が与えられており、系統樹上の距離に基づく、系統樹ガウス過程によるモデルを提案する。

データとしては、日本におけるカミキリムシの植物上における分布データを用いる。数百種類の全国に分布する植物上において、数百種類のカミキリムシの観測状況が二値変数として与えられている。また、植物の系統樹情報も得られている。カミキリムシはその生態から、系統樹上の距離が近い植物上に分布しやすいことが予測できるため、系統樹上の距離が近い植物上では、似た分布パターンを示すことが予測できる。一方で、カミキリムシの種間の分布パターンにおける類似性に関しては、事前知識を仮定せず、モデルの枠組みで類似性を捉えたい。

提案モデル

$S = \{s_1, \dots, s_n\}$ を観測された位置の集合とし、 $U_i := U(s_i) \in \mathbb{R}^S$ を s_i における $S \times 1$ の連続潜在ベクトルとする。Taylor-Rodríguez et al. (2016) は種間の依存構造をとらえるモデルとして因子モデルを提案した。

$$U_i = \mathbf{B}x_i + \mathbf{A}w_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}_S(\mathbf{0}, \sigma_c^2 \mathbf{I}_S), \quad \text{for } i = 1, \dots, n \quad (1)$$

ただし、 $w_i \sim \mathcal{N}_r(\mathbf{0}, \mathbf{I}_r)$ 。 $r \ll S$ とし、パラメータ数を減らしている。また、 \mathbf{A} の各行をディリクレ過程より生成することにより、種間の依存構造のクラスタリングを捉えるモデルとなっている。Shirota et al. (2019) では、 w に空間ガウス過程を導入し、予測のパフォーマンス向上を示した。本提案では、 w に系統樹情報より得られた距離を考慮したガウス過程を導入し、予測のパフォーマンスが向上することを示す。

参考文献

- Shirota, S., A. E. Gelfand, and S. Banerjee (2019). Spatial joint species distribution modeling using dirichlet processes. *Statistica Sinica*. DOI:10.5705/ss.202017.0482.
- Taylor-Rodríguez, D., K. Kaufeld, E. M. Schliep, J. S. Clark, and A. E. Gelfand (2016). Joint species distribution modeling: dimension reduction using Dirichlet processes. *Bayesian Analysis*.