Poisson_main

May 14, 2020

\#
Multiscale Geographically Weighted Regression - Poisson dependent variable
The model has been explored and tested for multiple parameters on real and simulated datasets. The research includes the following outline with separate notebooks for each part.

Notebook Outline:

Back to the main page

---

1 Introduction

1.1 Introduction to the problem

A recent addition to the local statistical models in PySAL is the implementation of Multiscale Geographically Weighted Regression (MGWR) model, a multiscale extension to the widely used approach for modeling process spatial heterogeneity - Geographically Weighted Regression (GWR). GWR is a local spatial multivariate statistical modeling technique embedded within the regression framework that is calibrated and estimates covariate parameters at each location using borrowed data from neighboring observations. The extent of neighboring observations used for calibration is interpreted as the indicator of scale for the spatial processes and is assumed to be constant across covariates in GWR. MGWR, using a back-fitting algorithm relaxes the assumption that all processes being modeled operate at the same spatial scale and estimates a unique indicator of scale for each process. The GWR model in PySAL can currently estimate Gaussian, Poisson and Logistic models though the MGWR model is currently limited to only Gaussian models. This project aims to expand the MGWR model to nonlinear local spatial regression modeling techniques where the response outcomes may be discrete (following a Poisson distribution). This will enable a richer and holistic local statistical modeling framework to model multi-scale process heterogeneity for the open source community.

1.2 Statistical Equations

A conventional Poisson regression model is written as:

$$O_i \sim Poisson[E_i exp(\sum \beta_k x_{k,i})] \tag{1}$$

$$\tag{2}$$

where $x_{k,1}$ is the kth explanatory variable in place i and the $\beta_k s$ are the parameters and Poisson indicates a Poisson distribution with mean λ.

Nakaya et.al. (2005) introduced the concept of allowing parameter values to vary with geographical location (u_i), which is a vector of two dimensional co-ordinates describing the location i. The Poisson model for geographically varying parameters can be written as:

$$O_i \sim Poisson[E_i exp(\sum \beta_k(u_i) x_{k,i})] \tag{3}$$

$$\tag{4}$$

The Geographically Weighted Poisson Regression model (GWPR) is estimated using a modified local Fisher scoring procedure, a form of iteratively reweighted least squares (IRLS). In this procedure, the following matrix computation of weighted least squares should be repeated to update parameter estimates until they converge (Nakaya et.al., 2005):

$$\beta^{(l+1)}(u_i) = (X^t W(u_i) A(u_i)^{(l)} X)^{-1} X^t W(u_i) A(u_i)^{(l)} z(u_i)(l) \tag{5}$$

$$\tag{6}$$

2 Approaches Explored

Expected theoretical model calibration: The calibration methodology for modeling response variables with a Poisson distribution in MGWR, through references from Geographically Weighted Poisson Regression (GWPR) (Nakaya, et al., 2005) and literature on Generalized Additive Models (Hastie & Tibshirani, 1986), is expected to be as follows:

1. Initialize using GWPR estimates $_()^0$: f1_0,f2_0, ... , f^_0, $ where $f^0 = (_1(_1)^0()^0)$
2. Update for each location () an adjusted dependent variable: $z()^{()} = (z_1()^{()}, z_2()^{()},, z()^{()})$ $z()^{()} = ()^{()}, + OO(()^{()})/O(()^{()})$
3. Construct weights as follows: $A()^{()}$ which is a diagonal matrix with values $(O_1(()^{()}), O_2(()^{()}), ..., O(()^{()}))$
4. Fit an MGWR model to $z()^{()}$ to update () and $z()^{()}$, using the new weight matrix: $()^{()} = ()^{()} A()^{()}$
5. Repeat steps (2) through (4) until convergence.

3 Other notebooks

Below are links to the tests and exploration for the finalized MGWR model for Poisson dependent variables:

3.1 Tests with real and simulated data

Initial module changes and univariate model check - Setup with libraries - Fundamental equations for Poisson MGWR - Example Dataset - Helper functions - Univariate example - Parameter check - Bandwidths check

Simulated Data example - Setup with libraries - Create Simulated Dataset - Forming independent variables - Creating y variable with Poisson distribution - Univariate example - Bandwidth: Random initialization check - Parameters check - Multivariate example - Bandwidths: Random initialization check - Parameters check - Global model parameter check

Real Data example

- Setup with libraries
- Tokyo Mortality Dataset
- Univariate example

    – Bandwidth: Random initialization check
    – Parameter check

- Multivariate example Bandwidths: Random initialization check
- MGWR bandwidths
- AIC, AICc, BIC check

3.2 Monte Carlo Test

Monte Carlo Simulation Visualization

- Setup with libraries
- List bandwidths from pickles
- Parameter functions
- GWR bandwidth
- MGWR bandwidths
- AIC, AICc, BIC check

    – AIC, AICc, BIC Boxplots for comparison

- Parameter comparison from MGWR and GWR

4 References

1. Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale Geographically Weighted Regression (MGWR). Annals of the American Association of Geographers, 107(6), 1247–1265. https://doi.org/10.1080/24694452.2017.1352480

2. Nakaya, T., Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2005). Geographically weighted Poisson regression for disease association mapping. Statistics in Medicine, 24(17), 2695–2717. https://doi.org/10.1002/sim.2129

3. Yu, H., Fotheringham, A. S., Li, Z., Oshan, T., Kang, W., & Wolf, L. J. (2019). Inference in Multiscale Geographically Weighted Regression. Geographical Analysis, gean.12189. https://doi.org/10.1111/gean.12189

4. Hastie, T., & Tibshirani, R. (1986). Generalized Additive Models. Statistical Science, 1(3), 297–310. https://doi.org/10.1214/ss/1177013604

5. Wood, S. N. (2006). Generalized additive models : an introduction with R. Chapman & Hall/CRC.

Back to the main page