Ataskaita už 2020/2021 doktorantūros metus

Doktorantė: Neringa Urbonaitė
Vadovas: Prof. habil. dr. Leonidas Sakalauskas

Vilniaus Universitetas
Duomenų mokslo ir skaitmeninių technologijų institutas

2019-2023

Disertacijos tema

Disertacijos tema:

Fraktalinio Brauno Lauko tyrimas ir taikymas daugiamačių duomenų modeliavime

Vadovas:

Prof. habil. dr. Leonidas Sakalauskas

Studijų metai	Egzaminai		Konferencijos		Publikacijos
-	Planas	Įvykdyta	Planas	Įvykdyta	Planas
I 2019-2020	2	2	1	1	–
II 2020-2021	2	1	1	–	–
III 2021-2022	–	–	1	–	1
IV 2022-2023	–	–	–	–	1

lentelė 1: Disertacijos rengimo planas

Einamieji studijų metai (II: 2020/2021).

Egzaminai		Konferencijos		Publikacijos	
Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta
Fundamentalieji informatikos ir inžierijos metodai	Išlaikyta	-	-	—	—
Gilieji neuroniniai tinklai	-	ASMDA 2021 liepos 1- 4 d.	Registracija	—	—

lentelė 2: Disertacijos rengimo planas

Mokslinių tyrimų ir disertacijos rengimo etapai

Darbo pavadinimas		Atlikimo terminai	Pastabos
1.	Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje): 1. Atlikti daugiamačių Krigingo modelių apžvalgą. 2. Nustatyti (identifikuoti) mokslines problemas, kylančias uždaviniuose.	2019 m. spalio mėn. – 2020 m. rugsėjo mėn.	~18psl.
2.	Mokslinio tyrimo vykdymas: 2.1. Tyrimo metodikos sudarymas; 2.1.1Tyrimo metodikos iškeltam uždaviniui spręsti parinkimas; 2.1.2 Teorinio ir empirinio tyrimų suplanavimas pagal pasirinktą metodiką. 2.2. Teorinis tyrimas: 2.2.1 Fraktalinių Brauno laukų (FBL) savybių tyrimas; 2.2.2 FBL realizacijų kompiuterinio imitavimo metodo sukūrimas. 2.3. Empirinis tyrimas: 2.3.1 FBL pritaikymas daugiamačiam krigingui ir ekstrepoliavimui; 2.3.2 Atstikitinių laukų parametrų vertinimo metodų: semivariogramos ir didžiausio tikėtinumo algoritmų sukūrimas. 2.3.3 Kompiuterinių eksperimentų atlikimas	2020 m. spalio mėn. – 2021 m. balandžio mėn. 2020 m. gegužės mėn. – 2021 m. rugsėjo mėn. 2021 m. lapkričio mėn. – 2022 m. rugsėjo mėn.	~17psl. ~23-24psl.

Disertacijos uždaviniai

Tyrimo objektas: Fraktalinis vektorinis Brauno laukas

Tikslas: Fraktalinių vektorinių Brauno laukų savybių tyrimas ir taikymas daugiamačiams duomenims modeliuoti.

Uždaviniai:

1. Sukurti FVBL'o modelį;
2. Sudaryti FVBL'o realizacijų generavimo algoritmą;
3. Sudaryti FVBL'o vertinimo algoritmą taikant didžiausio tikėtinumo metodą (DT) ir palyginti jį su variogramos (V) metodu;
4. Sudaryti metodą daugiamačių duomenų ekstrepoliavimui taikant FVBL'o modelį;
5. Pritaikyti sukurtą metodą praktiniams uždaviniams.

Yra zinomi du Brauno laukų formalizavimo būdai pagal Kolmogorovo įvestą kovariacijų funkciją ir remiantis spektriniu proceso aprašymu.

Apibrėžimas pagal A. Kolmogorovą

Gauso atsitiktinis laukas $Z(X)$ kurio kovariacinė funkcija

$$C(x_{(1)}, x_{(2)}) = \frac{1}{2}(\mid x_{(1)} \mid^{2H} + \mid x_{(2)} \mid^{2H} - \mid x_{(1)} - x_{(2)} \mid^{2H}) \quad (1)$$

su Hursto $H \in\ (0, 1)$ parametru.

Fraktalinio vektorinio Brauno lauko modelis

Tarkime, turime duomenų rinkinį: $X = (x_{(1)}, x_{(2)}, ..., x_{(k)})$; $x_{(i)} \in \mathbb{R}^d$; $1 \leq i, j \leq k$; $\mathbf{1} = (1, 1, ..., 1)$, $A = A_{i,j}$

$$A = ((x_i - x_j)^T (x_i - x_j))^H \tag{2}$$

$$B_H(X) = 2\frac{\mathbf{1}\mathbf{1}^T}{\mathbf{1}^T A^{-1} \mathbf{1}} - A \tag{3}$$

H - Hursto indeksas

Taikant darbe formulę (3) nereikia žinoti koordinačių pradžios, kitaip nei tradiciniame apibrėžime pagal A. Kolmogorovą.

[3] N. Pozniak, L. Sakalauskas ir L. Saltyte, Kriging model with fractional euclidean distance matrices," Informatica, t. 30, pp. 367-390, 2019.

Tarkime, turime k taškų: $X = (x_{(1)}, x_{(2)}, ..., x_{(k)})$;
$x_{(i)} \in \mathbb{R}^d$; $1 \leq i \leq k$;
Dažniausiai tai yra dvimačiai taškai $d = 2$ aprašomi dekarto ar geografinėmis koordinatėmis (erdviniai duomenys).

Kiekvienoje šių erdvės lokacijų atlikus stebėjimus gaunamos reikšmės.

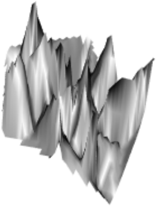
$Z = (z_{(1)}, z_{(2)}, ..., z_{(m)})$; $z_{(i)} \in \mathbb{R}^m$; $1 \leq i \leq k$.

Hursto indeksas

Žinant daugiamačių duomenų rinkiniui būdingą Hursto parametrą galima nuspėti chaotines duomenų išsidėstymo savybes, o vėliau atlikti prognozę ir interpoliaciją.

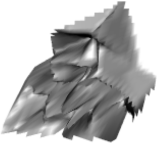
- Hursto parametras $H \in \ (0, 1)$.
- Dažniausiai pasitaikanti problema, Hursto indeksas nėra žinomas, todėl ieškoma metodų kurie šį parametrą gali ¿ivertinti.
- Pasirinkti du metodai ¿ivertinti Hursto parametrą: didžiausio tikėtinumo ir variogramos.

Fraktalinis vektorinis Brauno laukas

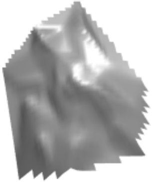


pav. 1: $H = 0.1$

pav. 2: $H = 0.9$



pav. 3: $H = 0.1$



pav. 4: $H = 0.9$

Variograma

Vienas iš metodų įvertinti parametrą H yra taikant variogramą. Tai fundamentalus modelis skirtas erdvinių porcesų modeliavimui.

Geostatistikos darbuose žinoma, kad variograma iš atsitiktinės funkcijos gali būti suskaidyta į atskirus taškus.

$$V_{(1)} = \frac{2}{k(k-1)}(\sum_{i=1}^{k-1}\sum_{j=1}^{k}(z_i - z_j)(z_i - z_j))$$

$$V_{(2)} = \frac{\sum_{ii=1}^{m}\sum_{jj=1}^{m}(\sum_{i=1}^{k-1}\sum_{j=1}^{k}(z_{ii,i}-z_{ii,j})(z_{jj,i}-z_{jj,j})((x_i-x_j)^T(x_i-x_j))^H)^2}{\sum_{i=1}^{k}\sum_{j=1}^{k}((x_i-x_j)^T(x_i-x_j))^{2H}}$$

$$V(H) = V_{(1)} - V_{(2)} \tag{4}$$

Iš formulės (6) optimali parametro reikšmė įvertinama minimizuojant pagal H.

Didžiausio tikėtinumo metodas

Taikant Didžiausio tikėtinumo metodą įvertinamos tokios parametrų reikšmės, su kuriomis gaunami rezultatai yra labiausiai tikėtini duotajam modeliui.

$$L = \frac{1}{K}(Z\ B_H(X)^{-1}Z^T - \frac{(Z\ B_H(X)^{-1}\mathbf{1})}{\mathbf{1}^T\ B_H(X)^{-1}\mathbf{1}}) \tag{5}$$

$$G(H) = \frac{1}{m}ln(L) + \frac{1}{k}ln(B_H(X)) \tag{6}$$

Tradiciškai yra taikomas variogramos metodas, o didžiausio tikėtinumo metodo taikymas erdvinių duomenų analizei yra mažai tyrinėtas.
Didžiausio tiketinumo metodą bandoma dažniausiai sudaryti tik procesams per Kolmogorovo tipo kovariacijų funkciją.

Monte Karlo statistinis eksperimentas

Eksperimento procedūra

1 žingsnis Parenkamas Hursto parametras $H = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

2 žingsnis Fraktalinio Brauno Lauko realizacijų generavimas. Trys duomenų grupės, po 100, kai $k = 20; 50; 100$ (eilučių skaičius). Lokacijų stulpelių buvo du $m = 2$, požymių stulpeliai $p = $ 2;5;10;.

3 žingsnis Taikant DT metodą ir variogramos metodą ¿vertinamas H parametras.

4 žingsnis Eksperimento procedūra kartojama.

5 žingsnis Gautiems H ¿iverčiams skaičiuojami vidurkiai η ir standartiniai nuokrypiai σ.

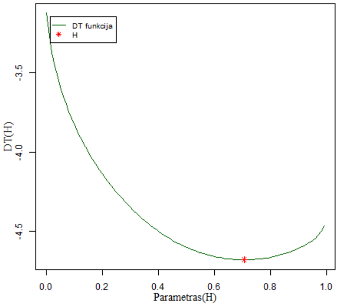
Choleckio metodas

Choleckio algoritmas remiasi FVBL kovariacijų matricos Cholecky dekompozicija.

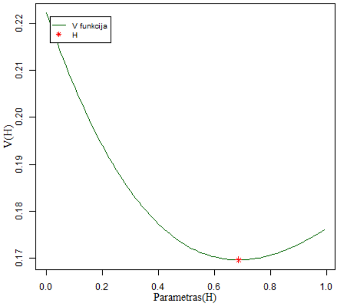
Tikslas: generuoti Fraktalinį Brauno Lauką su iš anksto parinktu Hursto parametru.

- Įvedama simetrinė matrica b; $\beta = chol(b)$;
- Generuojamas atsitiktinių dydžių vektorius ϑ;
- $C = chol(B_H(X))$;
- Paskaičiuojama FBVL realizacija $\beta\vartheta C^T$;

Atlikus perskaičiavimus gaunama FVBL realizacijų matrica Z(x$_{(i)}$),1<i<k

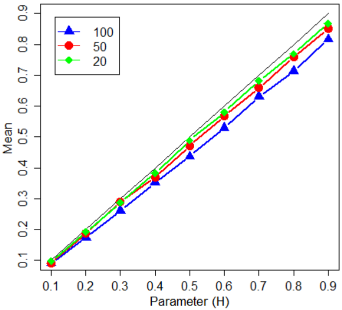


pav. 5: DT funkcija (minimali reikšmė $H = 0.7$)

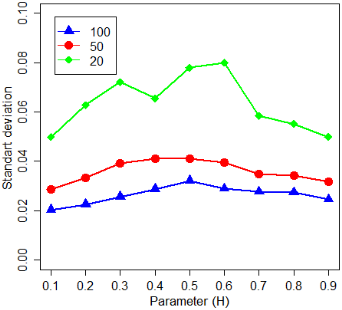


pav. 6: V funkcija (minimali reikšmė $H = 0.7$)

Monte Karlo statistinis eksperimentas: DT metodas

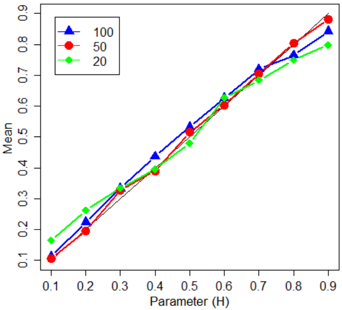


pav. 7: Vidurkiai įvertintų parametrų H

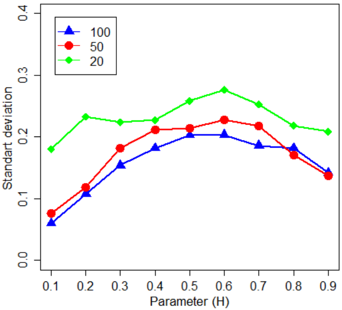


pav. 8: Standartiniai nuokrypiai parametrų H

Monte Karlo statistinis eksperimentas: V metodas



pav. 9: Vidurkiai įvertintų parametrų H



pav. 10: Standartiniai nuokrypiai parametrų H

Praktinis taikymas

Pasirinkti skirtingų metų (2017, 2018, 2019, 2020), kiekvieno mėnesio Lietuvos teritorijos kritulių ir temperatūros duomenis (Global Climate Monitor).

pav. 11: Krituliai (liepos mėn.)

pav. 12: Temperatūra

mėn.	2017	2018	2019	2020
1	0.84	0.81	0.83	0.83
2	0.81	0.82	0.80	0.81
3	0.78	0.73	0.71	0.71
4	0.67	0.72	0.70	0.75
5	0.75	0.69	0.65	0.74
6	0.72	0.77	0.72	0.66
7	0.70	0.76	0.65	0.71
8	0.63	0.67	0.70	
9	0.60	0.67	0.59	
10	0.66	0.65	0.64	
11	0.73	0.76	0.71	
12	0.72	0.83	0.79	

lentelė 3: Rezultatai, Hursto ¿iverčiai

Krigingas

Tarkime, turime duomenų rinkinį: $X = (x_{(1)}, x_{(2)}, ..., x_{(k)})$; $x_{(i)} \in \mathbb{R}^d$; $1 \leq i, j \leq k$; $\mathbf{1} = (1, 1, ..., 1)$, $A = A_{i,j}$

$$A = ((x_i - x_j)^T (x_i - x_j))^H \tag{7}$$

$$y_A(X) = z^T A^{-1} (b + 1 \frac{\mathbf{1} - \mathbf{1}^T A^{-1} b}{\mathbf{1}^T A^{-1} \mathbf{1}}) \tag{8}$$

H - Hursto indeksas

[3] N. Pozniak, L. Sakalauskas ir L. Saltyte, Kriging model with fractional euclidean distance matrices,“ Informatica, t. 30, pp. 367-390, 2019.

2020/2021 m. m. darbo planas

Disertacijos rengimo planas

1. Metodų taikymas praktiniams uždaviniams;
2. Išlaikyti egzaminą (1);
3. Sudalyvauti konferencijoje;
4. Parengti publikaciją;

Ačiū!