

Ataskaita už pirmus doktorantūros metus

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

Vilniaus Universitetas
Duomenų mokslo ir skaitmeninių technologijų institutas

2019-2023

Disertacijos tema:

Fraktalinių Brauno Laukų tyrimas ir taikymas daugiamačių duomenų modeliavimui

Vadovas:

Prof. habil. dr. Leonidas Sakalauskas

Doktorantūros pradžios ir pabaigos metai:

2019-2023

Pirmų metų disertacijos rengimo planas

- 1 Išlaikyti egzaminus (2);
- 2 Sudalyvauti bendruosiu gebėjimus stiprinančiose veiklose;
- 3 Atlikti mokslinių tyrimų disertacijos tema apžvalgą;
- 4 Sudaryti tyrimo metodiką;
- 5 Atlikti Fraktalinių vektorinių Brauno laukų (FVBL) savybių tyrimą;

Egzaminai išlaikyti per pirmus doktorantūros metus.

Dalykas	Kreditai	Laik. data	Komisija	Įvert.
Statistinis modeliavimas ir stochastinis optimizavimas	7	2019 gruodis	...	8
Informatikos ir informatikos inžinerijos tyrimo metodai ir metodikos	8	2020 birželis	...	10

- 1 SeThi 2020 - Online (Rostokas, Vokietija) Summer School
"Security of Things" 2020 Rugsjūtis 24 - Rugsėjis 4 (2ECTS)
- 2 Tampere Summer School 2020 - Online (Tampere, Suomija)
Introduction to Bioinformatics 2020 Rugsjūtis (2ECTS)
- 3 DAMSS 2019 (Druskininkai, Lietuva)

- 1 Atvirosios prieigos kompetencijų tobulinimas, taikant žaidimo metodą, 2019 gruodžio 18 d., (0.5 ECTS)
- 2 Informacijos tvarkymas programa Mendeley: praktinis užeiėmimas, 2019 gruodžio 3 d., (0.15 ECTS)
- 3 Mokslinė informacija: paieška, mokslometrija, duomenų talpyklos, 2019 lapkričio 13 d., (0.25 ECTS)

Tyrimo objektas:

Fraktalinis vektorinis Brauno laukas

Tyrimo tikslas:

Sukurti fraktalinio vektorinio Brauno lauko matematinį modelį naudojant euklido atstumų matricą. Sudaryti šio modelio parametro identifikavimo didžiausio tikėtimumo metodu algoritmus ir pritaikyti sukurtą modelį prognozavimo bei atpažinimo uždaviniams spręsti.

- 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.

- 1 Pasiektas rezultatas FvBI'o realizacijų generavimas Choleckio ir sąlyginių vidurkių metodais.
- 2 FvBI'o įidentifikavimo didžiausio tikėtimumo metodu algoritmas.
- 3 Daugiamačių duomenų ekstrapoliavimo taikant FvBI'o modelį algoritmas.

Atsitiktinis išsibarstymas yra būdingas daugiamatiams duomenims, kurie gali būti modeliuojami taikant **Fraktalinį vektorinį Brauno lauko modelį**.

Šis modelis pasižymi koreliacija nusakoma indeksu (H) dar dažnai vadinamu Hursto parametru $H \in (0, 1)$.

Tarkime, turime k taškų: $X = (x_{(1)}, x_{(2)}, \dots, x_{(k)});$

$$x_{(i)} \in \mathbb{R}^d; 1 \leq i \leq k;$$

Dažniausiai tai yra dvimačiai taškai $d = 2$ aprašomi dekartu ar geografinėmis koordinatėmis (erdviniai duomenys).

Kiekvienoje šių erdvės lokacijų atlikus stebėjimus gaunamos reikšmės.

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

- Brauno judesys (standartinis Vynerio procesas);
- Fraktalinis Brauno judesys (fBj) (1940 Andrėjus Kolmogorovas);
- B. Mandelbrot ir van Ness;
- Dauguma tyrimų yra atlikta kai Brauno laukas yra vienmatis (N. Pozniak);

Yra žinomi 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

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

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

Spektrinis metodas gana abstraktus juo yra pažengta sukuriant formalią teoriją kuri daugiausiai aprašo fraktalinio lauko realizacijų elgesį.

[1]I. S. Reed, P. C. Lee, T. K. Truong, Spectral representation of Fractional Brownian Motion in n dimensions and its properties, Transactions on Information Theory, nr. 5, 1995

Dažnai duomenims fizikinę prigimtį sunku atsekti, todėl taikomi metodai, kurie remiasi įvairiomis skaitmeninėmis paradigmomis. Vienas iš taikomų modelių krigingas [2].

[2] N. Zhang, D. W. Apley; Fractional Brownian fields for response surface metamodeling; Journal of Quality Technology; Vol. 46, Issue 4; (2014)

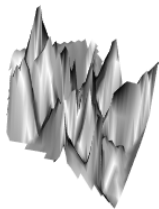
Šiame darbe yra nagrinėjama vektorinio fraktalinio lauko formalizacija per atstumų matricą (stebejimo taškų) remiantis N.Pozniak darbu [3].

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

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 įvertinti.
- Pasirinkti du metodai įvertinti Hursto parametrą: didžiausio tikėtimumo ir variogramos.

Fraktalinis vektorinis Brauno laukas



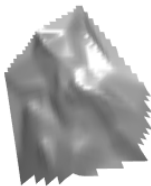
pav. 1: $H = 0.1$



pav. 2: $H = 0.9$



pav. 3: $H = 0.1$



pav. 4: $H = 0.9$

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

$$A = ((x_i - x_j)^T (x_i - x_j))^H \quad (2)$$

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

H - Hursto indeksas

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

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 $Z(x_{(i)})$,
 $1 \leq i \leq k$

Vienas iš metodų įvertinti parametą 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.

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

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

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

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

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

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

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

Tradiciškai yra taikomas variogramos metodas, o didžiausio tikėtinumo metodo taikymas erdvinių duomenų analizei yra mažai tyrinėtas.

Didžiausio tikėtinumo metodą bandoma dažniausiai sudaryti tik procesams per Kolmogorovo tipo kovariacijų funkciją.

Eksperimento procedūra

1 žingsnis Parenkamas Hursto parametras

$H = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

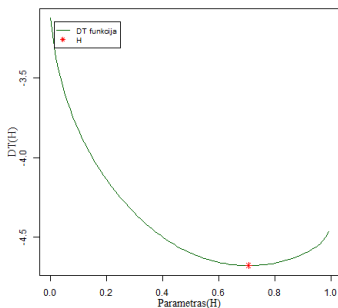
2 žingsnis FVBL'o realizacijų generavimas. Trys duomenų grupės, po 100, kai $k = 20; 50; 100$ (eilučių skaičius). Lokacijų stulpelių buvo du $m = 2$, požymių stulpeliai $p = 2; 5; 10$;

3 žingsnis Taikant DT metodą ir variogramos metodą įvertinamas H parametras.

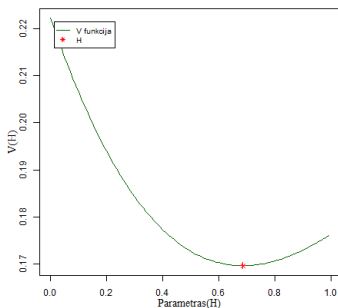
4 žingsnis Eksperimento procedūra kartojama.

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

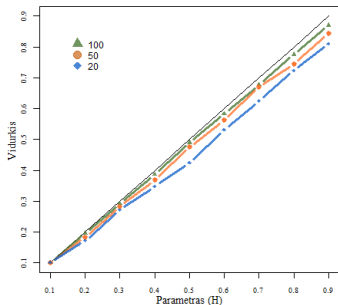
Monte Karlo statistinis eksperimentas ($H = 0.7$)



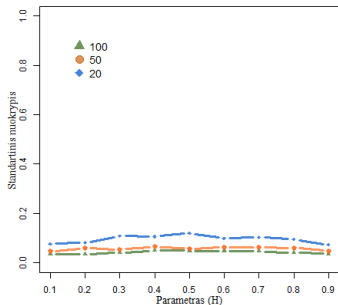
pav. 5: DT funkcija (minimali reikšmė $H = 0.7$)



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

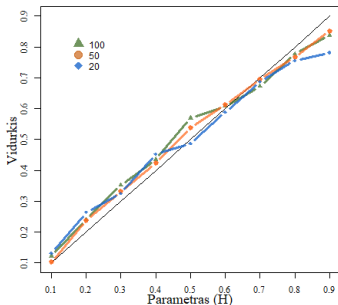


pav. 7: Vidurkiai įvertintų parametrų H

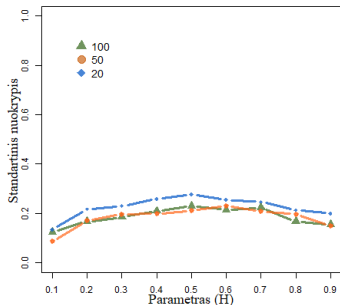


pav. 8: Standartiniai nuokrypiai parametrų H

Monte Karlo statistinis eksperimentas: \forall metodas



pav. 9: Vidurkiai įvertintų parametrų H



pav. 10: Standartiniai nuokrypiai parametrų H

Antrų metų disertacijos rengimo planas

- 1 Metodų taikymas praktiniams uždaviniams;
Pasirinkti skirtingų metų (2017, 2018, 2019, 2020), kiekvieno mėnesio Lietuvos teritorijos kritulių ir temperatūros duomenys (Global Climate Monitor).
- 2 Išlaikyti egzaminus (2);
- 3 Sudalyvauti konferencijoje;

Ačiū!