



VILNIAUS UNIVERSITETAS  
MATEMATIKOS IR INFORMATIKOS INSTITUTAS

**DOKTORANTŪROS METINĖ ATASKAITA**  
2016 m. spalio mėn. 1 d. – 2017 m. rugsėjo mėn. 30 d.

INFORMATIKOS STUDIJŲ PROGRAMOS  
**DOKTORANTĖ JŪRATĖ VAIČIULYTĖ**

- **DISERTACIJOS TEMA:**  
PASLĖPTŪJŲ MARKOVO MODELIŲ TAIKYMAS APTARNAVIMO  
SYSTEMŲ BEI TINKLŲ MODELIAVIMUI IR OPTIMIZAVIMUI
- **VADOVAS PROF. HABIL. DR. LEONIDAS SAKALAIUSKAS**

- Įstojimo į doktorantūrą metai 2015 m. spalio mėn. 1 d.
- Doktorantūros baigimo metai 2019 m. rugsėjo mėn. 30 d.

- **Tyrimo objektas**
  - Paslėptieji Markovo modeliai ir taikymas
- **Tyrimo tikslas**
  - Paslėptųjų Markovo modelių parametrų įvertinimo rekurentinių algoritmų tyrimas
- **Tyrimo uždaviniai**
  - Analitiškai apžvelgti paslėptųjų Markovo modelių parametrų vertinimo metodus
  - Sukurti metodus ir pritaikyti juos praktiniams uždaviniams spręsti
- **Planuojami rezultatai**
  - Sukurtų metodų pritaikymas praktiniams uždaviniams spręsti (šnekos atpažinimui) ir gautų rezultatų taikymo ne Gauso modeliams apibendrinimas ir algoritmai.

# 2016/2017 m. m. darbo planas

- Išlaikyti egzaminus: Duomenų gavyba; Daugiamatė statistika, laiko eilutės
- Mokslinių tyrimų planas:
  - Tyrimo metodikos iškeltam uždaviniui spręsti parinkimas
  - Paslėptųjų Markovo modelio kalibravimo metodų tyrimas
- Dalyvauti mokslinėje konferencijoje
- Parengti 1 publikaciją.

# 2016/2017 m. m. atlikti darbai

- Egzaminas „Duomenų gavyba“ išlaikytas
- Egzaminas „Daugiamatė statistika, laiko eilutės“ nebuvo laikytas (yra įteiktas vadovo prašymas dėl dalyko pakeitimo).

# 2016/2017 m. m. atlikti darbai

- Atliktas Paslėptųjų Markovo modelių kalibravimo metodų tyrimas
  - Sudarytas rekurentinis paslėptųjų Markovo modelių parametru vertinimo algoritmas.
  - Atlikus Monte-Karlo eksperimentus sudarytu algoritmu nustatyta, kad sudaryto rekurentinio algoritmo sudėtingumas yra pirmos eilės ir leidžia atlikti kalibravimą realiu laiku.
  - Sudarytas rekurentinis algoritmas pritaikytas daugiamačių duomenų klasterizavimui.
  - Atliktas sudaryto algoritmo efektyvumo tyrimas statistinio modeliavimo būdu.

# 2016/2017 m. m. atlikti darbai

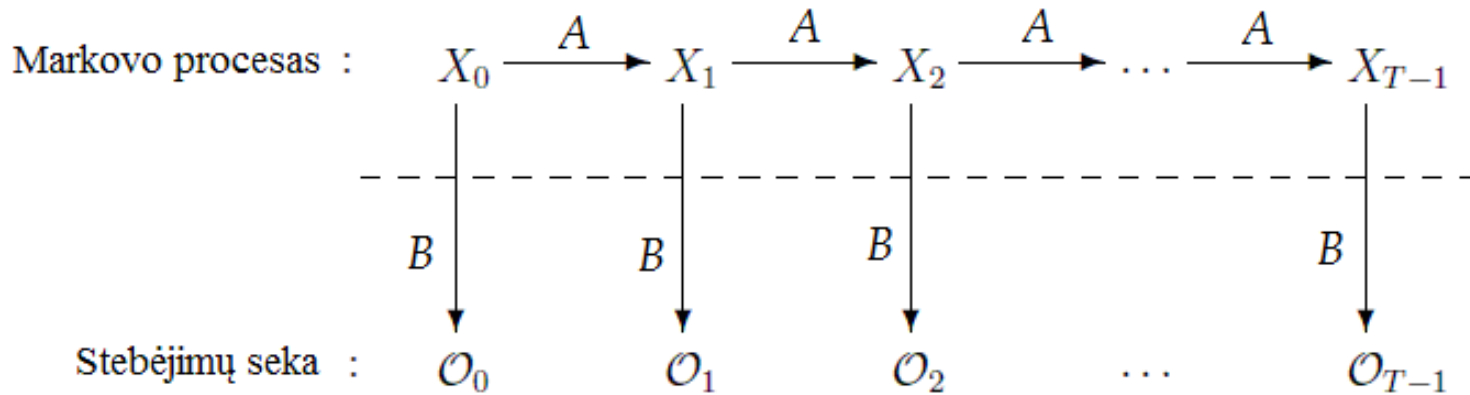
- Dalyvauta respublikinėje mokslinėje-praktinėje konferencijoje „Informacinių technologijų iššūkiai kūrybos ekonomikoje“.
- Dalyvauta „Kompiuterininkų dienų 2017“ konferencijoje. Pranešimo tema: “Rekurentinis paslėptųjų Markovo modelių parametro vertinimo algoritmas“.
- Dalyvauta tarptautinėje 8-ojoje „Data Analysis Methods for Software Systems“ konferencijoje. Pranešimo tema: “Recurrent estimation of homogeneous Hidden Markov model parameters“.
- Dalyvauta tarptautinėje vasaros mokykloje „27th Jyväskylä Summer School“ Suomijoje.



# 2016/2017 m. m. atlikti darbai

- 1 publikacija priimta spausdinimui žurnale „Computational Science and Techniques“. Publikacijos tema: “ Rekurentinis paslėptųjų Markovo modelių parametro vertinimo algoritmas”
- 1 publikacija įteikta recenzavimui žurnale „Pattern Recognition“. Tema: „Recurrent estimation of multivariate hidden Markov model parameters“

# PMM matematinio modelio parametrai



- Būsenų skaičius  $N$
- būsenų perėjimų tikimybių matrica  $A = \begin{pmatrix} a_{11} & \dots & a_{1N} \\ \vdots & \dots & \vdots \\ a_{N1} & \dots & a_{NN} \end{pmatrix}$
- pradinio buvimo būsenoje tikimybių vektorius  $\pi = \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_N \end{pmatrix}$
- stebėjimo tikimybės tankio funkcija

$$N(o, \mu_s, \sigma_s) = \frac{1}{\sqrt{(2\pi)^n |\sigma_s|}} \exp\left(-\frac{1}{2}(o - \mu_s)^T \sigma_s^{-1} (o - \mu_s)\right)$$

- stebėjimus aprašantys atsitiktiniai dydžiai, pasiskirstę pagal duotus vidurkius  $\mu_s$  ir kovariacijų matricas  $\sigma_s, 1 \leq s \leq N$

# EM algoritmas

- Pasinaudojus didžiausio tikėtimumo metodu išvestos formulės PMM parametų vertinimui

$$\bar{\mu}_j = \frac{\sum_{j=1}^T \gamma(j) \cdot o}{\sum_{j=1}^T \gamma(j)} \quad \bar{\sigma}_j = \frac{\sum_{j=1}^T \gamma(j) \cdot (o - \bar{\mu}_j)(o - \bar{\mu}_j)^T}{\sum_{j=1}^T \gamma(j)} \quad \gamma_j = e^{-l(o, \bar{\mu}_j, \bar{\sigma}_j, \pi_j)}$$

$$l(o, \mu, \sigma, \pi) = \frac{(o - \mu)^T \sigma^{-1} (o - \mu)}{2} + \ln \left( \frac{\sqrt{|\sigma|}}{\pi} \right)$$

# Rekurentinis EM algoritmas

- Rekurentinės formulės vidurkių vektoriaus ir kovariacijų matricos įvertinimui:

$$\mu_t^i = \mu_{t-1}^i + \frac{(O_t - \mu_{t-1}^i) \cdot \omega_t^i}{t} \quad \sigma_t^i = \left( \frac{\beta_{t-1}^i \cdot (t-1)}{\beta_t^i \cdot t} \right) \cdot \left( \sigma_{t-1}^i + \frac{(O_t - \mu_{t-1}^i)(O_t - \mu_{t-1}^i)^T \cdot \omega_t^i}{t} \right)$$

- Būsenų perėjimo tikimybių skaičiavimo formulė:

$$\beta_t^i = \beta_{t-1}^i + \frac{1}{t} (\omega_t^i - \beta_{t-1}^i)$$

kur koeficientas  $\omega$ :

$$\omega_t^i = \frac{e^{-l(o_t, \hat{\mu}_i, \hat{\sigma}_i, \pi_i)}}{\gamma_t} \quad \gamma_t = \sum_{i=1}^N e^{-l(o_t, \hat{\mu}_i, \hat{\sigma}_i, \pi_i)}$$

# Eksperimento eiga. PMM parametrų vertinimas

- Sudarytam algoritmui patikrinti buvo atliktas kompiuterinis eksperimentas: daugiamačių stebėjimo vektorių klasterizavimas.
- Duoti daugiamačiai Gauso stebėjimų vektoriai, kurie sudaro du klasterius ( $k=2$ ). Klasterių tikrieji centrai:  $[500,500,\dots]$  ir  $[600,600,\dots]$ . Dimensijų skaičius: 2, 4, 8, 16. Standartinis nuokrypis: 20.
- Stebėjimo vektorių imtis: nuo 1000 iki 10000. Skaičiavimai atlikti su kiekviena imtimi po 100 kartų.
- Mokymui skirta fiksuoto dydžio stebėjimų vektorių imtis  $T=500$ .
- Parametrų vertinimui fiksuojamas algoritmų (rekurentinio EM ir klasikinio EM) vykdymo greitis sekundėmis.

# Eksperimento eiga. PMM parametru vertinimas

	M = 2		M = 4		M = 8		M = 16	
T=	Recurrent (time)	Batch (time)	Recurrent (time)	Batch (time)	Recurrent (time)	Batch (time)	Recurrent (time)	Batch (time)
1000	3,90	6,87	6,53	21,91	2,30	17,04	3,35	5,60
2000	4,05	13,74	6,56	27,23	2,64	23,49	3,59	9,30
3000	4,35	13,95	6,89	25,93	2,96	16,85	3,91	10,80
4000	4,72	13,95	7,20	28,07	3,29	15,72	4,23	17,51
5000	5,15	14,03	7,61	28,36	3,63	16,22	4,67	17,66
6000	5,70	15,26	8,09	30,17	4,04	18,53	5,10	26,39
7000	6,33	16,87	8,67	34,17	4,50	21,02	5,56	24,60
8000	7,01	19,27	9,36	38,43	5,09	23,18	6,15	35,79
9000	7,88	21,66	10,14	38,77	5,71	26,45	6,72	31,21
10000	8,75	24,06	10,99	43,96	6,46	55,51	7,35	44,33

## 2017/2018 m. m. darbo planas:

- Išlaikyti egzaminą.
- Mokslinių tyrimų planas:
  - Sukurtų paslėptųjų Markovo modelių eksperimentinis tyrimas analizuojant keletą testinių duomenų aibių;
  - Praktinių sričių, kuriose būtų tinkamas sukurtas paslėptųjų Markovo modelių metodas, nustatymas bei jo pritaikymas praktiniam uždaviniui.
  - Gautų rezultatų analizė, apibendrinimas, išvadų parengimas.
- Dalyvauti mokslinėje konferencijoje
- Parengti 1 publikaciją



VILNIAUS UNIVERSITETAS  
MATEMATIKOS IR INFORMATIKOS INSTITUTAS

**DOKTORANTŪROS METINĖ ATASKAITA**  
2016 m. spalio mėn. 1 d. – 2017 m. rugsėjo mėn. 30 d.

INFORMATIKOS STUDIJŲ PROGRAMOS  
**DOKTORANTĖ JŪRATĖ VAIČIULYTĖ**