



VILNIAUS UNIVERSITETO DUOMENŲ MOKSLO IR
SKAITMENINIŲ TECHNOLOGIJŲ INSTITUTAS

DOKTORANTŪROS METINĖ ATASKAITA

2017 m. spalio mėn. 1 d. – 2018 m. rugsėjo mėn. 30 d.

INFORMATIKOS STUDIJŲ PROGRAMOS
DOKTORANTĖ JŪRATĖ VAIČIULYTĖ

- **DISERTACIJOS TEMA:**
PASLĖPTŪJŲ DAUGIAMAČIŲ MARKOVO MODELIŲ TYRIMAS IR
TAIKYMAS REALAUS LAIKO SISTEMŲ ANALIZEI
- **VADOVAS PROF. HABIL. DR. LEONIDAS SAKALAUSKAS**

- 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 rekursyvinių algoritmų tyrimas ir taikymas
- **Tyrimo uždaviniai**
 - Analitiškai apžvelgti paslėptųjų Markovo modelių parametrų vertinimo metodus
 - Sukurti Gauso ir ne Gauso paslėptųjų Markovo modelių daugiamačių parametrų vertinimo algoritmus;
 - Sukurtus algoritmus iširti statistinio modeliavimo būdu ir palyginti su kitais PMM parametrų vertinimo algoritmais;
 - Sukurtus metodus pritaikyti praktiniams uždaviniams spręsti.
- **Planuojami rezultatai**
 - Sukurtų metodų pritaikymas praktiniams uždaviniams spręsti (klasteriavimui, izoliuotų objektų atpažinimui, šnekos atpažinimui) ir gautų rezultatų taikymo Gauso ir ne Gauso modeliams apibendrinimas ir algoritmai.

2017/2018 m. m. darbo planas

- Išlaikyti egzaminą „Atpažinimo teorija“.
- Mokslinių tyrimų planas:
 - Sukurtų paslėptųjų Markovo modelių eksperimentinis tyrimas analizuojant testinių duomenų aibes;
 - Praktinių sričių, kuriose būtų tinkamas sukurtas paslėptųjų Markovo modelių metodas, nustatymas bei jo pritaikymas praktiniams uždaviniams.
 - Gautų rezultatų analizė.
- Dalyvauti mokslinėje konferencijoje.
- Parengti 1 publikaciją.

2017/2018 m. m. atlikti darbai

- Egzaminas „Atpažinimo teorija“ išlaikytas
 - Egzamino komisija:
 - prof. habil. dr. Laimutis Telksnys
 - dr. Gražina Korvel
 - doc. dr. Gintautas Tamulevičius
 - Įvertinimas: 10 (puikiai)

2017/2018 m. m. atlikti darbai

- 1 publikacija įteikta recenzavimui žurnale „Computational Statistics“.
 - Tema: „Recursive Estimation of Multivariate Hidden Markov Model Parameters“

2017/2018 m. m. atlikti darbai

- Dalyvauta Lietuvos Matematikų Draugijos 59-oje konferencijoje. Pranešimo tema: "Rekurentinis paslėptųjų Markovo modelių parametru vertinimo algoritmas ir jo taikymai". 2018 m. birželio mėn. 18-19 d., Kaunas.
- Dalyvauta tarptautinėje konferencijoje EURO 2018 (29th European Conference on Operational Research). Pranešimo tema: „Recurrent parameter estimation algorithm in hidden Markov models with application to multivariate data analysis and signal recognition“. 2018 m. Liepos mėn. 8-11 d., Valencija.

Sudarytas rekursyvinis EM algoritmas PMM parametrų vertinimui

- Būsenų perėjimo tikimybės apskaičiuojamos pagal Chapman-Kolmogorov lygtis (pakeista klasikinė Forward-Backward procedūra, kuri įprastai rekursyviuose algoritmuose yra praleidžiama):

$$\pi_t = A \cdot \pi_{t-1}$$

- Vidurkių vektorius, kovariacijų matrica ir buvimo būsenoje tikimybės apskaičiuojamos pagal formules:

$$\mu_t^i = \mu_{t-1}^i + \frac{(o_t - \mu_{t-1}^i)}{t} \cdot \frac{\omega_t^i}{\beta_t^i},$$

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

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

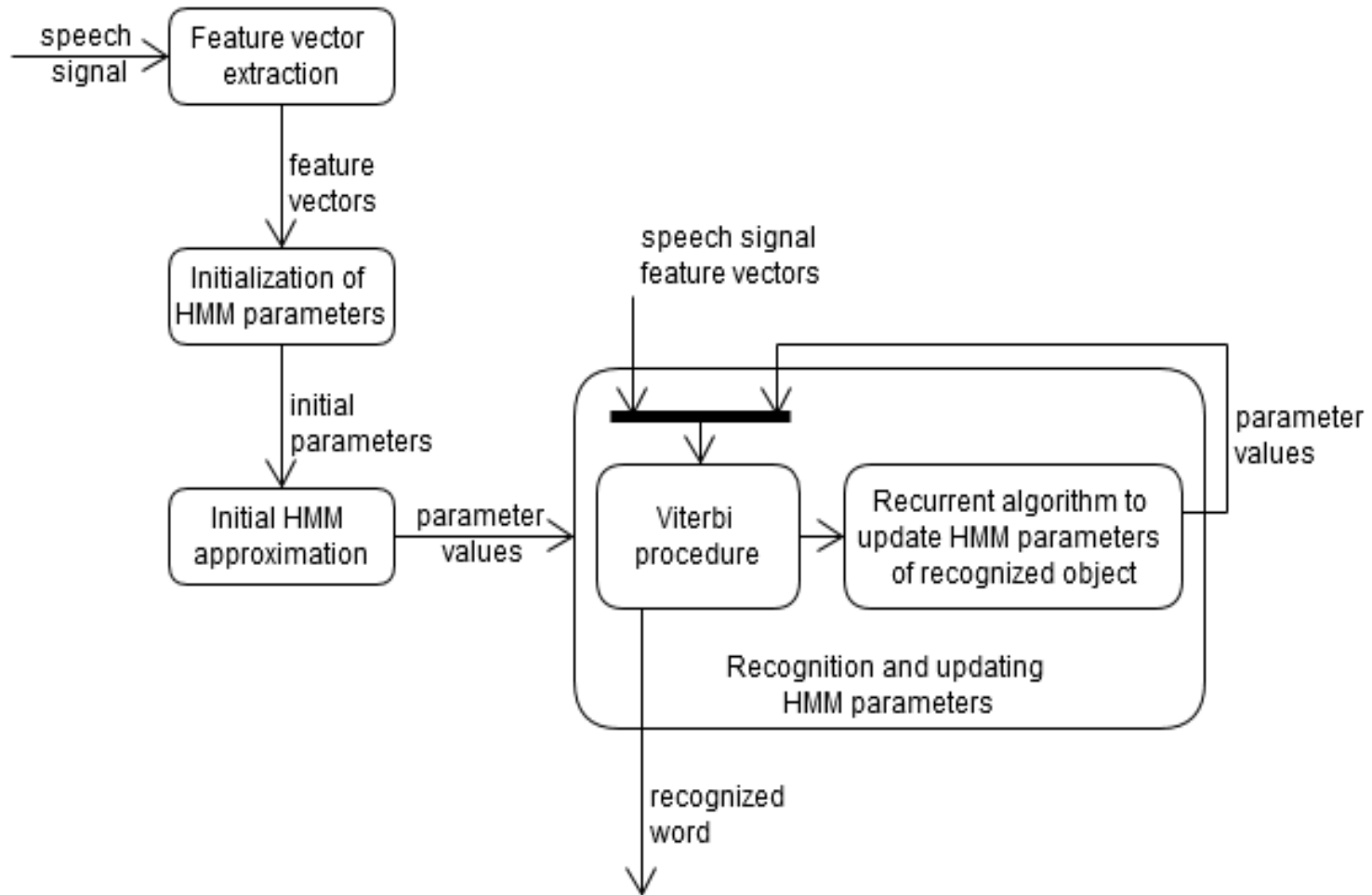
- čia koeficientai:

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

Logaritminė tikėtinumo f-ja:

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

PMM parametru vertinimo rekursyvinis EM algoritmas pritaikytas izoliuotų žodžių atpažinimui



Eksperimento rezultatai

- Eksperimentai atlikti su TIDIGITS izoliuotų žodžių duomenų aibe:

Size (in words) of the initial training dataset	Size (in words) of the re-estimation dataset	Recognition rate (%)
100	1500	89.27
500		94.53
700		94.33
1000		94.87
1500		95.00
1800		94.73
2000		94.58

Dirichlė PMM modelio parametrų vertinimas

- Dirichlė PMM parametrai:
 - PMM būsenų skaičius: N
 - Būsenų perėjimų tikimybių matrica
 - Pradinio buvimo būsenoje tikimybių vektorius
 - $\alpha_1, \dots, \alpha_K$, čia $\alpha_i > 0$ ir $K > 2$
 - x_1, \dots, x_K , čia $x_i \in (0,1)$ ir $\sum_{i=1}^K x_i = 1$
 - Logaritminė tikėtinumų funkcija:

$$\log Dir(x, \alpha) = \ln \Gamma \left(\sum_{i=1}^K \alpha_i \right) - \sum_{i=1}^K \ln \Gamma(\alpha_i) + \left[(\alpha_K - 1) \ln \left(1 - \sum_{i=1}^{K-1} x_i \right) \right] + \sum_{i=1}^{K-1} [(\alpha_i - 1) \ln(x_i)]$$

Didžiausio tikėtimumo metodu išvestos rekursyvinės formulės Dirichlet PMM parametrų vertinimui

- $\theta_t^{\langle q \rangle} = P_t^{\langle q \rangle} \log \text{Dir}(x_t, \alpha^{\langle q \rangle})$
- $\omega_t^{\langle q, s \rangle} = \omega_{t-1}^{\langle q, s \rangle} + \frac{1}{t} \left(\frac{\ln(x_t^{\langle s \rangle}) \theta_t^{\langle q \rangle}}{\sum_{j=1}^N \theta_t^{\langle j \rangle}} - \omega_{t-1}^{\langle q, s \rangle} \right), 1 \leq q < N$
- $\omega_t^{\langle q, K \rangle} = \omega_{t-1}^{\langle q, K \rangle} + \frac{1}{t} \left(\frac{\ln(1 - \sum_{s=1}^{K-1} x_t^{\langle s \rangle}) \theta_t^{\langle q \rangle}}{\sum_{j=1}^N \theta_t^{\langle j \rangle}} - \omega_{t-1}^{\langle q, K \rangle} \right)$
- $\gamma_t^{\langle q \rangle} = \gamma_{t-1}^{\langle q \rangle} + \frac{1}{t} \left(\frac{\theta_t^{\langle q \rangle}}{\sum_{j=1}^N \theta_t^{\langle j \rangle}} - \gamma_{t-1}^{\langle q \rangle} \right)$
- $\alpha_t^{\langle q, s \rangle} = \Psi^{-1} \left[\text{Psi} \left(\sum_{i=1}^K \alpha_i^{\langle q \rangle} \right) + \frac{\omega_t^{\langle q, s \rangle}}{\gamma_t^{\langle q \rangle}} \right], 1 \leq s \leq K$
 – Ψ^{-1} – atvirkštinė digamma funkcija

Atliktas eksperimentinis sudaryto algoritmo tyrimas statistinio modeliavimo būdu:

- Atliktas parametų vertinimas ir atpažinimas rekursiniais Dirichlet ir Gauso EM parametų vertinimo algoritmais.
- Generuojamos dvi stebėjimų vektorių aibės (vienos stebėjimų vektoriai pasiskirstę pagal Dirichlet skirstinį, kitos – pagal Gauso skirstinį)
- Sugeneruotų abiejų aibių tūris : $T=1000$;
- Eksperimentai pakartoti 100 kartų, iš kurių suskaičiuotas atpažinimo tikslumo vidurkis.
- Pradinė aproksimacija (apmokymas) atlikta su 200 stebėjimo vektorių.
- Paslėptųjų Markovo modelių parametrai:
 - Būsenų skaičius: 2;
 - Būsenų perėjimų tikimybių matrica: $P := \begin{pmatrix} 0.5 & 0.5 \\ 0.6 & 0.4 \end{pmatrix}$
 - Pradinių buvimo būsenoje tikimybių vektorius: $\pi := \begin{pmatrix} 0.7 \\ 0.3 \end{pmatrix}$
 - Stebėjimų vektorių dimensijų kiekis: $n=3$;

Atlikto eksperimento metu gauti atpažinimo tikslumo rezultatai

- Generuojami stebėjimų vektoriai, pasiskirstę pagal Dirichlet skirstinį:
 - *Gauso parametry vertinimas ir stebėjimų atpažinimas*
 - Atpažinimo tikslumo vidurkis: 52.8%
 - *Dirichlet parametry vertinimas ir stebėjimų atpažinimas*
 - Atpažinimo tikslumo vidurkis: 92.8%
- Generuojami stebėjimų vektoriai, pasiskirstę pagal Gauso skirstinį:
 - *Gauso parametry vertinimas ir stebėjimų atpažinimas*
 - Atpažinimo tikslumo vidurkis: 99.48%
 - *Dirichlet parametry vertinimas ir stebėjimų atpažinimas*
 - Atpažinimo tikslumo vidurkis: 95.2%

2018/2019 m. m. darbo planas:

- Parengti daktaro disertaciją (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų ir kt.).
- Dalyvauti mokslinėje konferencijoje.
- Parengti 1 publikaciją.



VILNIAUS UNIVERSITETO DUOMENŲ MOKSLO IR
SKAITMENINIŲ TECHNOLOGIJŲ INSTITUTAS

DOKTORANTŪROS METINĖ ATASKAITA

2017 m. spalio mėn. 1 d. – 2018 m. rugsėjo mėn. 30 d.

INFORMATIKOS STUDIJŲ PROGRAMOS
DOKTORANTĖ JŪRATĖ VAIČIULYTĖ