

DOKTORANTŪROS STUDIJŲ DALYKO SANDAS

Dalyko pavadinimas	Mokslo kryptis (šaka) kodas	Fakultetas	Katedra
Natūralios kalbos apdorojimas	Informatikos inžinerija, 07 T	MIF	Duomenų mokslo ir skaitmeninių technologijų institutas

Studijų būdas	Kreditų skaičius ECTS	Studijų būdas	Kreditų skaičius
paskaitos	1 (pavasario semestras)	konsultacijos	1
individualus	4	seminarai	1

Dalyko anotacija

Pageidautina, kad doktorantas pasirenkantis šį dalyką išmanytų, tiesinę algebrą, matematinę analizę ir Python programavimo kalbą.

Pagrindinis sando tikslas – gilinti natūralios kalbos apdorojimo žinias naudojant pažangius natūralios kalbos metodus ir technologijas.

Dalyko temos

1. Įvadas į natūralios kalbos apdorojimą. Pagrindinės sąvokos. Šiuolaikinių naudojamų sričių apžvalga. Lingvistiniai natūralios kalbos apdorojimo ištekčiai. Garsyno taksonomija.
2. Lingvistiniai kalbos elementai. Morfologija. Sintaksės analizė. Semantinė interpretacija. Kalbos pragmatika.
3. Žodžių atvaizdavimas į vektorius. Paprastieji atvaizdavimo metodai (word2vec, GloVe).
4. Pažangūs žodžių atvaizdavimo į vektorius metodai (kalbos modeliai (angl. skip-gram, bag of words), vieno žodžio klasifikavimas, logistinė regresija, viena-sluoksnis neuroninis tinklas).
5. Žodžių lango klasifikavimas ir neuroniniai tinklai. Neuroniniai tinklai, gradiento sklidimo atgal algoritmas, neuroninio tinklo persimokymas, reguliavimas, aktyvacijos funkcijos. naudojimas atpažįstant įvardintas esybes (angl. named entity recognition). Neuroninis tinklo modelis gramatinių frazių klasifikavimui. Tensorflow programinė įranga.
6. Rekurentiniai neuroniniai tinklai kalbos modeliavimui. Tradiciniai kalbos modeliai. Rekurentiniai neuroniniai tinklai. Dvikrypčiai rekurentiniai tinklai. Rekurentiniai neuroniniais tinklais grįsti kalbos modeliai jų mokymo problemos ir sprendimai. Nuomonių tyryba su rekurentiniais tinklais.
7. Rekurentiniai neuroniniai tinklai mašininiam vertimui. Statistiniai mašininio vertimo metodai. GRU (angl. Gated Recurrent Unit) ir LSTM (angl. Long Short-Term Memory) metodai mašininiam vertimui.
8. Rekursiniai neuroniniai tinklai. Rekursiniai neuroniniai tinklai struktūros prognozavimui – sintaksinei, gramatiniam ir semantiniam nagrinėjimui.
9. Pažangūs rekursiniai tinklai. Rekursiniai neuroniniai tinklai (RNT): standartiniai, Matrica-vektorius RNT, Rekursiniai neuroniniai tensorių tinklai, LSTM medžiai ir jų naudojimas atpasakojimo aptikimui, sąryšių klasifikavimui, sentimentų analizei, frazių panašumo nustatymui.

Praktinės užduotys:

1. Klasikinių algoritmų taikymas. Žodžių atvaizdavimas į vektorius arba žodžių lango klasifikavimas.
2. Specializuotų algoritmų kūrimas. Pagal doktoranto vykdomą mokslinę temą parinkta užduotis, kurios sprendimui būtų naudojami rekursiniai ar rekurentiniai neuroniniai tinklai.

Sandą sudarys 9 paskaitos, 9 seminarai, du atsiskaitomieji darbai.

Pagrindinė literatūra
Steven Bird, Ewan Klein, and Edward Loper Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit, 2015, http://www.nltk.org/book
Daniel Jurafsky and James Martin (2017). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (Third edition), Prentice-Hall, Inc., 1024 p. https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf
Alexander Clark, Chris Fox and Shalom Lappin (2013). The Handbook of Computational Linguistics and Natural Language Processing, John Wiley & Sons, 802 p.
Mohamed Zakaria Kurdi (2016). Natural Language Processing and Computational Linguistics: Speech, Morphology and Syntax. Wiley-ISTE, 296 p.
CS224d: Deep Learning for Natural Language Processing, http://cs224d.stanford.edu/syllabus.html

Konsultuojančiųjų dėstytojų vardas, pavardė	Mokslo laipsnis	Svarbiausieji darbai mokslo kryptyje (šakoje) paskelbti per pastaruosius 5 metus
Virginijus Marcinkevičius	dr.	<p>Medvedev V.; Kurasova O.; Bernatavičienė J.; Treigys P., Marcinkevičius V. Dzemyda G., 2017. A new web-based solution for modelling data mining processes. Simulation Modelling Practice and Theory. ISSN: 1569-190X, Vol. 74, Elsevier</p> <p>Jakštys, V.; Marcinkevičius, V.; Tichonov, J.; Treigys, P., 2016. Detection of the road pothole contour in raster images // Information technology and control. Kaunas: Technologija. ISSN: 1392-124X. 2016, Vol. 45, no. 3, p. 300-307.</p> <p>Bernatavičienė, J.; Dzemyda, G.; Bazilevičius, G.; Medvedev, V.; Marcinkevičius, V.; Treigys, P. 2015. Method for visual detection of similarities in medical streaming data // International journal of computers communications & control. Oradea: Universitatea Agora. ISSN 1841-9836. 2015, Vol. 10, no. 1, p. 8-21.</p>
Gražina Korvel	dr.	<p>G. Korvel, B. Kostek (2017). Voiceless stop consonant modelling and synthesis framework based on MISO dynamic system. Archives of Acoustics, 42 (3), 375-383.</p> <p>G. Korvel, V. Šimonytė, V. Slivinskas (2015). A modified additive synthesis method using source-filter model. Journal of the Audio Engineering Society, 63 (6), 443-450.</p> <p>G. Pyž, V. Šimonytė, V. Slivinskas (2014). Developing models of Lithuanian speech vowels and semivowels. Informatica 25 (1), 55-72</p>
Gintautas Tamulevičius	dr.	<p>Andrius Rybakovas, Virgilijus Beiša, Kęstutis Strupas, Jonas Kaukėnas, Gintautas Tamulevičius, „Inverse filtering of speech signal for detection of vocal fold paralysis after thyroidectomy“, Informatica, Vol. 29, No. 1, 2018, p. 91-105.</p> <p>Tatjana Liogienė, Gintautas Tamulevičius, „Multi-stage recognition of speech emotion using sequential forward feature selection“, Electrical, control and communication engineering, Vol. 10, No. 1, 2016, p. 35-41.</p> <p>Gintautas Tamulevičius, Tatjana Liogienė, „Low-order multi-level features for speech emotion recognition“, Baltic journal of modern computing, Vol. 3, No. 4, 2015, p. 234-247.</p> <p>Andrius Rybakovas, Virgilijus Beiša, Kęstutis Strupas, Jonas Kaukėnas, Gintautas Tamulevičius, „Inverse</p>

	filtering of speech signal for detection of vocal fold paralysis after thyroidectomy“, Informatica, Vol. 29, No. 1, 2018, p. 91-105.
--	--