

DOKTORANTŪROS STUDIJŲ DALYKO SANDAS

Dalyko pavadinimas	Mokslo kryptis (šaka) kodas	Fakultetas	Katedra
Didžiųjų duomenų analitika	Informatikos inžinerija, 07T	MIF	Duomenų mokslo ir skaitmeninių technologijų institutas

Studijų būdas	Kreditų skaičius ECTS	Studijų būdas	Kreditų skaičius
paskaitos	1 (rudens sem.)	konsultacijos	1
individualus	4	seminarai	1

Dalyko anotacija

Dalyko tikslas – siekiama, kad doktorantas įgytų esminių įgūdžių apie didžiųjų duomenų analitiką bei ugdytų praktinius gebėjimus taikyti didžiųjų duomenų saugojimo ir apdorojimo būdus ir technologijas bei tyrybos metodus.

Įvadinė dalis: Didžiųjų duomenų apibrėžtys bei jų vystymosi istorija. Pagrindiniai didžiųjų duomenų šaltiniai. Didžiųjų duomenų tipai: struktūriniai ir nestrukūriniai duomenys. Didžiųjų duomenų saugojimo būdai ir technologijos. Didžiųjų duomenų tyryba: taikymo uždavinių pavyzdžiai.

Didžiųjų duomenų analitika:

Duomenų tyryba žinių gavimo procese, žinių gavimo proceso etapai.

Didžiųjų duomenų pirminio apdorojimo būdai.

Didžiųjų duomenų analitikos uždaviniai

Didžiųjų duomenų tyryba:

- Panašumų ir dažnų sekų paieška
- Prognozavimo ir klasifikavimo metodai, paremti didžiųjų duomenų analize
- Srautinių duomenų analizė
- Klasterizavimo metodai
- Dimensijos mažinimo būdai
- Mašininis mokymasis didelių apimčių duomenims: neuroniniai tinklai ir gilusis mokymas
- Teksto tyrybos metodai.
- Didžiųjų duomenų vizualizavimo iššūkiai.

Didžiųjų duomenų technologijos: MapReduce paradigma. Apache Hadoop ir Spark platformos. Spark MLlib biblioteka mašininiam mokymuisi, Didžiųjų duomenų analitika taikant R paketą.

Debesų kompiuterijos sprendimai, taikomi didžiųjų duomenų analitikoje. Microsoft Azure Machine Learning platforma. Amazon Web Service sprendimai. KNIME Analytics Platform praplėtimai didiesiems duomenims.

Didžiųjų duomenų tyrybos praktiniai aspektai. Praktinių taikymų pavyzdžiai sprendžiant uždavinius, kylančius versle, finansuose, medicinoje ir kt.

Praktinė užduotis: išspręsti nurodytą duomenų analizės uždavinį, taikant didžiųjų duomenų technologijas.

Pagrindinė literatūra

Marr, Bernard. Big data : using smart big data, analytics and metrics to make better decisions and improve performance, 2015, ISBN 978-1-118-96583-2

Simon Walkowiak. Big Data Analytics with R, Publisher: Packt Publishing, 2016, ISBN: 9781786466457

Mining Massive Data Sets, SOE-YCS0007, Stanford School of Engineering

Konsultuojančiųjų dėstytojų vardas, pavardė	Mokslo laipsnis	Svarbiausieji darbai mokslo kryptyje (šakoje) paskelbti per pastaruosius 5 metus
Viktor Medvedev	dr.	<ul style="list-style-type: none"> • 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. Amsterdam: Elsevier Science. ISSN 1569-190X. eISSN 1878-1462. 2017, Vol. 76, p. 34-46. DOI: 10.1016/j.simpat.2017.03.001 • Bernatavičienė, J.; Dzemyda, G.; Kurasova, O.; Marcinkevičius, V.; Medvedev, V.; Treigys, P. Cloud computing approach for intelligent visualization of multidimensional data // Advances in stochastic and deterministic global optimization / Editors: Panos M. Pardalos, Anatoly Zhigljavsky, Julius Žilinskas. - Ser.: Springer optimization and its applications. ISSN 1931-6828. Vol. 107. Cham: Springer International Publishing Switzerland, 2016. ISBN 9783319299730. p. 73-85. DOI: 10.1007/978-3-319-29975-4_5. • Medvedev, V.; Kurasova, O. Cloud technologies: a new level for Big Data Mining // Resource management for Big Data platforms: algorithms, modelling, and high-performance computing techniques / Editors: Florin Pop, Joanna Kołodziej, Beniamino Di Martino. Ser.: Computer communications and networks. ISSN 1617-7975. Cham: Springer International Publishing AG, 2016. ISBN 9783319448800. p. 55-68. DOI: 10.1007/978-3-319-44881-7_3 • 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.
Jolita Bernatavičienė	dr.	<ul style="list-style-type: none"> • Morkūnas, M.; Treigys, P.; Bernatavičienė, J.; Laurinavičius, A.; Korvel, G.. Machine learning based classification of colorectal cancer tumour tissue in whole-slide images // Informatica. Vilnius : Vilniaus universiteto Matematikos ir informatikos institutas. ISSN 0868-4952. 2018, Vol. 29, no. 1, p. 75-90. DOI: 10.15388/Informatica.2018.158. • 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. Amsterdam: Elsevier Science. ISSN 1569-190X. eISSN 1878-1462. 2017, Vol. 76, p. 34-46. DOI: 10.1016/j.simpat.2017.03.001 • Bernatavičienė, J.; Dzemyda, G.; Kurasova, O.; Marcinkevičius, V.; Medvedev, V.; Treigys, P. Cloud computing approach for intelligent visualization of

		<p>multidimensional data // Advances in stochastic and deterministic global optimization / Editors: Panos M. Pardalos, Anatoly Zhigljavsky, Julius Žilinskas. - Ser.: Springer optimization and its applications. ISSN 1931-6828. Vol. 107. Cham: Springer International Publishing Switzerland, 2016. ISBN 9783319299730. p. 73-85. DOI: 10.1007/978-3-319-29975-4_5.</p> <ul style="list-style-type: none"> • 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 (ISI Web of Science. Scopus)
Gintautas Dzemyda	habil. dr.	<ul style="list-style-type: none"> • Gėgžna, V., O. Kurasova, G. Dzemyda, R. Kurtinaitienė, I. Čiplys, J. V. Vaitkus, A. Vaitkuvienė (2018). The ROC-based analysis of spectroscopic signals from medical specimens. Nonlinear Analysis: Modelling and Control, 23 (3), 285-302. • Bilinskas, M.J., Dzemyda, G., Trakymas, M. (2018). Approximation of the ribs-bounded contour in a Tomography scan slice. International Journal of Information Technology & Decision Making, 17 (1), 83-102. https://doi.org/10.1142/S0219622017500298 • Evora, J., J. J. Hernandez, M. Hernandez, G. Dzemyda, O. Kurasova, E. Kremers (2015). Swarm intelligence for frequency management in smart grids. Informatica, 26 (3), 419-434.