

ATTILA KERTÉSZ-FARKAS, PH.D.

DOCENT - ASSISTANT PROFESSOR

FACULTY OF COMPUTER SCIENCE
NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS (HSE)
MOSCOW, RUSSIA

<http://www.hse.ru/en/staff/akerteszfarkas>
akerteszfarkas@hse.ru

AREA OF EXPERTISE

Machine Learning, Data mining, and Bioinformatics.

EDUCATION

- 2010 Ph.D. in Computer Science, University of Szeged, Hungary, *summa cum laude*
Thesis: "Protein Classification in a Machine Learning Framework"
Advisors: Prof. János Csirik, Dr. András Kocsor
- 2004 M.Sc. in Computer Science and Mathematics, University of Szeged, Hungary, *summa cum laude*
Thesis: "Compact Representation of Finite Languages with Nondeterministic Finite Automata"
Advisor: Prof. Zoltán Fülöp

RESEARCH EXPERIENCES

- 2015-present Assistant Professor (Docent), School of Data Analysis and Artificial Intelligence, Faculty of Computer Science, National Research University Higher School of Economics (HSE), Moscow, Russian Federation
- 2013-2015 Postdoctoral Fellow, Department of Genome Sciences, University of Washington, Seattle WA, USA
- 2009-2013 Postdoctoral Fellow, Bioinformatics Group, International Centre of Genetic Engineering and Biotechnology (ICGEB), Trieste, Italy
- 2008-2009 Research Fellow, Division of Imaging and Applied Mathematics, CDRH, U.S. Food and Drug Administration (U.S. FDA), Silver Spring MD, USA. Joint affiliation with Department of Biology, University of Maryland Baltimore County (UMBC), Catonsville MD, USA
- 2004-2008 Ph.D. Student, Doctoral School of Computer Science, University of Szeged, Hungary
- 2000-2004 Undergraduate Research Assistant, Research Group on Artificial Intelligence, Hungarian Academy of Sciences, Szeged, Hungary

SCHOLARSHIPS AND AWARDS

- 2007-2008 Predoctoral scholarship, awarded by the Foundation for Research and Teaching of Informatics, Szeged, Hungary
- 2006-2007 Erasmus scholarship, winter semester at Technische Universität Dresden, Germany
- 2003-2008 Honors College (Eötvös Collegium) for outstanding students at University of Szeged, Hungary
- 2004 1st prize at a Scientific Conference of Students for the article entitled "Kernel-based learning with feature weighting", University of Szeged
- 2002 1st prize at a Scientific Conference of Students for the article entitled "Compact representation of finite languages with nondeterministic automata", University of Szeged

TEACHING EXPERIENCES

- 2015-present Lecturer of Advanced Topics in Machine Learning (in English), HSE, Moscow, Russia
- 2015 Lecturer at Bioinformatics Tools for Life Sciences Workshop (in English), Quito, Ecuador
- 2012 Lecturer at Theoretical and Practical Course on Advanced Bioinformatics (Eng.), ICGEB, Italy
- 2010-2013 Teaching assistant, Theoretical and Practical Course on Bioinformatics (Eng.), ICGEB, Italy
- 2007 Teaching assistant, Operational Research, University of Szeged
- 2006 Teaching assistant, Algorithms and Data Structures, University of Szeged
- 2004-2005 Teaching assistant Introduction to Pascal, C and Java Programming, University of Szeged

MEMBERSHIPS

- 2014-present Industrial/Research Chair, Biotechno conference series
- 2014-present Editorial board of Applied and Computational Mathematics, Science Publication Group
- 2012-present Editorial board of International Journal on Advances in Life Sciences
- 2011-2015 Consultant at ManuSTAT central (a statistical consulting service for life sciences)
- 2006-2014 John von Neumann Computer Science Society, Hungary
- 2005-2007 PhD student representative on the Board of Doctoral Schools, Faculty of Sci., Univ of Szeged
- 2005-2007 PhD student representative on the Board of the Institute of Informatics, University of Szeged

LIST OF PUBLICATIONS

Number of independent citations to my publications is over 260.

THESES

1. Ph.D.: “*Protein Classification in a Machine Learning Framework*” 2008
2. M.Sc.: “*Compact Representation of Finite Languages with Nondeterministic Finite Automata*” 2004

BOOK CHAPTERS

3. Attila Kertész-Farkas, András Kocsor and Sándor Pongor: *The application of Data Compression-based Distance to Biological Sequences*; In: Frank Emmert-Streib. (Ed.) Information Theory and Statistical Learning, 73-88, 2008
4. Dmitry V. Nikitin, Attila Kertész-Farkas, Alexander S. Solonin and Marina L. Mokrishcheva: *Bifunctional prokaryotic DNA-methyltransferases*; In: Dr. Anica Dricu (Ed.) Methylation (ISBN 980-953-307-146-1), 2012

PEER-REVIEWED JOURNAL PAPERS (*equally contributed authors; †corresponding author)

5. Uri Keich, Attila Kertész-Farkas, and William Stafford Noble: *Mix-Max: an improved false discovery rate estimation procedure for shotgun proteomics*, Journal of Proteome Research, 14(8) 3148-3161, 2015
6. Attila Kertész-Farkas, Uri Keich, and William Stafford Noble: *Tandem mass spectrum identification via Cascaded search*; Journal of Proteome Research, 14(8) 3027-3038, 2015
7. Bruna Marini, Attila Kertész-Farkas, Hashim Ali, Bojana Lucic, Kamil Lisek, Lara Manganaro, Sándor Pongor, Roberto Luzzati, Alessandra Recchia, Fulvio Mavilio, Mauro Giacca, and Marina Lusic: *Nuclear architecture dictates HIV-1 integration site selection*; Nature 521, 227-231, 2015

8. Sean McIlwain, Kaipo Tamura, [Attila Kertész-Farkas](#), Charles E. Grant, Benjamin Diament, Barbara Frewen, J. Jeffry Howbert, Michael R. Hoopmann, Lukas Käll, Jimmy K. Eng, Michael J. MacCoss, and William Stafford Noble: *Crux: rapid open source protein tandem mass spectrometry analysis*; *Journal of Proteome Research* 13(10) 4488-4491, 2014
9. [Attila Kertész-Farkas](#), Beáta Reiz, Roberto Vera, Michael P. Myers, and Sándor Pongor: *PTMTreeSearch: a Novel Two-Stage Tree Search Algorithm with Pruning Rules for the Identification of Post-Translational Modification of Proteins in MS/MS Spectra*; *Bioinformatics* 30(2): 234-241, 2014
10. Dmitry V. Nikitin, [Attila Kertész-Farkas](#), Alexander S. Solonin, and Marina L Mokrishcheva: *Bifunctional restriction-modification enzymes: products of natural genetic engineering*; submitted, 2014
11. János Juhász*, [Attila Kertész-Farkas*](#), Dóra Szabo, and Sándor Pongor: *Emergence of Collective Territorial Defense in Bacterial Communities: Horizontal Gene Transfer Can Stabilize Microbiomes*; *PLoS One* 9(4), 2014
12. Roberto Vera, Yasset Perez-Riverol, Sonia Perez, Balázs Ligeti, [Attila Kertész-Farkas](#), and Sándor Pongor: *JBioWH: an open-source Java framework for bioinformatics data integration*; Database, 2013
13. Beáta Reiz, Michael P. Myers, Sándor Pongor, and [Attila Kertész-Farkas†](#): *Precursor Mass Dependent filtering of Mass Spectra for Proteomics Analysis*; *Protein and Peptide Letters* 21(8) 858-863, 2013
14. Beáta Reiz, [Attila Kertész-Farkas](#), Somdutta Dhir, Sándor Pongor, and Michael P. Myers: *Chemical rule-based filtering of MS/MS spectra*; *Bioinformatics* 29(7) 925-932, 2013
15. [Attila Kertész-Farkas](#), Beáta Reiz, Michael P. Myers, and Sándor Pongor: *Database searching in mass spectrometry based proteomics*; *Current Bioinformatics* 7(2) 221-230, 2012
16. Dóra Bihary, Ádám Kerényi, Zsolt Gelencsér, Sergiu Netotea, [Attila Kertész-Farkas](#), Vittorio Venturi, and Sándor Pongor: *Simulation of communication and cooperation in multispecies bacterial communities with an agent based model*; *Scalable Computing: Practice and Experience* 13(1) 21–28, 2012
17. Beáta Reiz, [Attila Kertész-Farkas](#), Sándor Pongor, and Michael P. Myers: *Data preprocessing and filtering in mass spectrometry based proteomics*; *Current Bioinformatics* 7(2) 212-220, 2012
18. [Attila Kertész-Farkas](#), Beáta Reiz, Michael P. Myres, and Sándor Pongor: *PTMSearch: A Greedy Tree Traversal Algorithm for Finding Protein Post-Translational Modifications in Tandem Mass Spectra*; In: *European Conference on Machine Learning and Principles and Practical Knowledge Discovery in Databases 2*, Vol. 6912 Springer, 162-176, 2011
19. Emily Doughty*, [Attila Kertész-Farkas*](#), Olivier Bodenreider, Gary Thompson, Asa Adadey, Thomas Peterson, and Maricel G. Kann: *Toward an automatic method for extracting cancer- and other disease-related point mutations from the biomedical literature*; *Bioinformatics* 27(3) 408-415, 2011
20. Somdutta Dhir, Mircea Pacurar, Dino Franklin, Zoltán Gáspári, [Attila Kertész-Farkas](#), András Kocsor, Frank Eisenhaber and Sándor Pongor: *Detecting atypical examples of known domain types by sequence similarity searching: The SBASE domain library approach*; *Current Protein Peptide Science* 11(7) 538-549, 2010
21. József Dombi and [Attila Kertész-Farkas†](#): *Applying Fuzzy Technologies to Equivalence Learning in Protein Classification*; *Journal of Computational Biology* 16(4) 611-623, 2009
22. Róbert Busa-Fekete, [Attila Kertész-Farkas](#), András Kocsor, and Sándor Pongor: *Balanced ROC analysis (BAROC) protocol for the evaluation of protein similarities*; *Journal of Biochemical and Biophysical Methods* 70(6) 1210-1214, 2008
23. [Attila Kertész-Farkas](#), Somdutta Dhir, Paolo Sonogo, Mircea Pacurar, Sergiu Netoteia, Harm Nijveen, Arnold Kuzniar, Jack A.M. Leunissen, András Kocsor, and Sándor Pongor: *Benchmarking protein classification algorithms via supervised cross-validation*; *Journal of Biochemical and Biophysical Methods* 70(6) 1215-1223, 2008

24. Attila Kertész-Farkas, András Kocsor, and Sándor Pongor: *Equivalence Learning in Protein Classification*; In: P. Perner (Ed.) *Machine Learning and Data Mining in Pattern Recognition*, Springer Verlag, Heidelberg, LNAI (4571) 824-837, 2007
25. János Z. Kelemen, Attila Kertész-Farkas, András Kocsor, and László G. Puskás: *Kalman Filtering for Disease-State Estimation from Microarray Data*; *Bioinformatics* 22(24) 3047-3053, 2006
26. László Kaján, Attila Kertész-Farkas, Dino Franklin, Nelly Ivanova, András Kocsor, and Sándor Pongor: *Application of a simple log likelihood ratio approximant to protein sequence classification*; *Bioinformatics* 22(23) 2865-2869, 2006
27. Paolo Sonogo, Mircea Pacurar, Somdutta Dhir, Attila Kertész-Farkas, András Kocsor, Zoltán Gáspári, Jack A.M. Leunissen, and Sándor Pongor: *A Protein Classification Benchmark collection for Machine Learning*; *Nucleic Acids Research* 35 232-236, 2006
28. András Kocsor, Attila Kertész-Farkas, László Kaján, and Sándor Pongor: *Application of compression-based distance measures to protein sequence-classification: a methodological study*; *Bioinformatics* 22(4) 407-412, 2006
29. Attila Kertész-Farkas and András Kocsor: *Kernel-based Classification of Tissues using Feature Weightings*; *Applied Ecology and Environmental Research* 4(2) 63-71, 2006
30. Attila Kertész-Farkas, Zoltán Fülöp, and András Kocsor: *Compact Representation of Hungarian Corpora*; (in Hungarian), *Hungarian Journal of Applied Linguistics* (1-2) 63-70, 2005

CONFERENCES (oral presentations)

31. *Peptide identification in tandem mass spectrometry data via cascade search*; US HUPO; Seattle USA, 6-9 April 2014,
32. *PTMSearch: A Greedy Tree Traversal Algorithm for Finding Protein Post-Translational Modifications in Tandem Mass Spectra*; European Conference on Machine Learning and Principles and Practice of knowledge discovery in databases; Athens Greece, 5-9 September 2011
33. *Equivalence Learning in Protein Classification*; International Conference on Machine Learning and Data Mining in Pattern Recognition; Leipzig Germany, 18-20 July 2007
34. *Classification of Tissues using Feature Weightings*; VII. Hungarian Conference on Biometrics and Biomathematics; Budapest Hungary, 5-6 July 2005.
35. *Kernel-based learning with dimension weighting*; Scientific Conference of Students, University of Szeged; April 2004,
36. *Compact Representation of Hungarian Vocabulary with Nondeterministic Finite Automata*; I. Conference on Hungarian Computational Linguistic; Szeged Hungary, 10-11 December 2003
37. *Compact representation of finite languages with nondeterministic automata*; Scientific Conference of Students, University of Szeged; October 2002,

CONFERENCES (poster presentation)

38. *EMU: A tool for the Extraction of Mutations with disease associations from literature*; Growth Factor and Signal Transduction Conferences: System Biology, Integrative, Comparative and Multi-scale Modeling, Iowa USA, 11-14 June 2009

TALKS

39. *Protein databases and similarity measures for protein classification*; Bioinformatics seminar at Rényi Mathematical Institute; Budapest Hungary, October 2007
40. *Equivalence Learning in Protein Classification*; Young Researcher Symposium on Intelligent Systems at the John von Neumann Computer Society; Budapest Hungary, November 2007

41. *A new algorithm for identification of post-translational modification in tandem mass spectrometry data*; Symposium at International Centre for Genetic Engineering and Biotechnology; Trieste Italy, June 2011
42. *Noise filtering and compression of mass spectra for proteomics*; Symposium at International Centre for Genetic Engineering and Biotechnology; Trieste Italy, June 2012
43. *Algorithmic Identification of Post-translational Modifications of Proteins in Tandem Mass Spectra*, Luxembourg Centre for Systems Biomedicine, Université du Luxembourg, Luxembourg, April 2013
44. *Statistical Methods in MS/MS spectrum identification*, Universidad San Francisco de Quito, Ecuador, October 2015