

# Dmitriy Fradkin

---

*URL:* <https://sites.google.com/site/dfradkin/>  
*Citizenship:* USA

*E-mail:* [dfradkin@gmail.com](mailto:dfradkin@gmail.com)  
*Location:* Wayne, PA

## RESEARCH INTERESTS

Exciting applications of data mining and machine learning, such as predictive maintenance, text classification, bioinformatics and search.

## EDUCATION

**Rutgers University**, New Brunswick, NJ

Ph.D, Computer Science - January 2006

GPA 3.793

Advisors: Casimir Kulikowski, Ilya Muchnik

Thesis Title: "Within-Class and Unsupervised Clustering Improve Accuracy and Extract Local Structure for Supervised Classification"

**Brandeis University**, Waltham, MA

B.A. in Computer Science, Mathematics, May 1999

Double Major, completed in three years with High Honors

Shapiro Award in Mathematics (May 1999)

Phi Beta Kappa

GPA 3.86

Dean's List every semester

## WORK EXPERIENCE

**Siemens Corporate Research**, Princeton, NJ

*Senior Scientist*

*Staff Scientist*

*Research Scientist*

**January 2015 - present**

**January 2012 - December 2014**

**June 2007 - December 2011**

Participated in multiple predictive maintenance, condition monitoring, text analytics and business intelligence projects, worked on development and implementation of tools for data integration and analysis, supervised junior colleagues and interns, wrote papers, invention disclosures, project and grant proposals, functional and test specifications, received PM@R&D certification.

- Performed analysis of gas turbine reliability (R, Oracle, Teradata, Spotfire) (2015-present)
- Developed and implemented algorithms for a Hadoop-based analytics platform (Java, Scala, Python, R, Spark, MongoDB, Tableau, Teradata) (2011-present)
- Worked on algorithms and developed tools for sequential pattern mining, log file, sensor and fleet-wide analysis of medical scanner data (Java, Matlab, MongoDB) (2009-2014)
- Performed data analysis on Next-Generation Sequencing data (SQL, R, Tableau) (2012-2013)
- Participated in development of a node of National Geothermal Database (Java, R, SQL), responsible for data import and integration (2011-2013)
- Developed a system for automated failure detection in MR machines (Java); supervised transfer of the algorithm to SAS (2008, 2012)
- Led development of prototype Natural Language Querying for Condition Monitoring Database prototype (2011)
- Participated in development of a data mining plug-in and multiple components for iKDD platform (Java) (2007-2010)
- Developed and evaluated methods for PowerSemantics project (Java) (2009)
- Improved speed and accuracy of a SPEND text classification system by utilizing sparse classifier models and spelling correction (2008, 2009)(Java)

- Developed a method for analysis of blood spectroscopy data (Java, Matlab) (2008)
- Analyzed gene expression data for biomarker detection and drug target identification, worked with SNP and rtPCR data (Matlab, R) (2007-2008)

**Ask.com**, Edison, NJ

*Senior Software Engineer*

**January 2006 - June 2007**

*Classification and Online Relevance groups*

- Worked on improving ranking and text scoring algorithms (Perl)
- Implemented automated parameter selection methods (Perl)
- Constructed a system for automatically identifying important urls (Perl)
- Assisted colleagues with development and use of algorithms for data analysis and classification

**Rutgers University**, New Brunswick, NJ

*Graduate Student*

**September 1999 - January 2006**

Ph.D. level research and coursework/projects in machine learning, pattern recognition, statistics, bioinformatics, image understanding, algorithms, linear programming.

- Developed a library of algorithms for K-Means clustering and a program for applying these algorithms (Java)
- Extended LIBSVM software to include backwards feature selection (C)
- Wrote programs for cross-validation, feature selection (Perl, shell script, R).
- Developed and implemented scripting systems for combining cluster analysis software with supervised classification (SVM, logistic regression) software (Perl, shell script, R).
- Wrote functions for image analysis (Matlab), stochastic optimization (Lisp)

*Graduate Assistant*

**September 2001 - December 2005**

Member of Monitoring Message Streams (MMS) and AuthorID Working Groups at DIMACS (Center for Discrete Mathematics and Theoretical Computer Science) (September 2002 - December 2005)

- Designed, implemented and evaluated methods for text classification, adaptive filtering, fusion of classifiers and interpolation of systems:
  - Developed servlets for visual information presentation (Java)
  - Participated in development of Adaptive Filtering Software package (C++)
  - Created programs for k-nearest-neighbor classification, local learning, fusion and interpolation of classification and information retrieval methods (C++,Perl, R, Octave)
- Wrote technical reports and high-level presentations

Worked on searching stochastically generated design spaces (September 2001 - August 2002)

- Developed Lisp programs for evaluating fitness of design populations produced by genetic programming.

*Teaching Assistant*

**September 1999 - May 2001**

Numerical Analysis and Computing (cs323, fall 2000 - spring 2001)

Data Structures (cs112, fall 1999 - spring 2000);

- Taught three recitation sections (over 60 students) each semester
- Presented material on Data Structures, Algorithms, Java programming and Numerical Methods
- Assisted students with assignments and course material
- Graded homework assignments and exams

- Substituted for lecturers

**Skila Inc.**, Mahwah, NJ

*Application Programmer*

**Summer 2000 & Summer 2001**

Participated in development of Skila System - intelligence system for Health Care Industry.

- Developed agents to acquire and process data in XML and HTML formats. Used Java, Microsoft Messaging Queue, Visual Parse++, Microsoft J++, Microsoft Visual Source Safe, SQL Navigator
- Developed HTML and ASP scripts for displaying results of SQL queries in thin client version and client-server version of Skila System (ASP, HTML, SQL, Oracle, Sybase SQL AnyWhere, SQL Navigator, source integrity system MKS).

#### COMPUTER SKILLS

- Languages: Java, Scala, Perl, Python, SQL, C, C++,
- Numerical Tools: Matlab, R, Octave
- Databases: Oracle, SQL Server, PosGres, Teradata, Mongo
- Big Data: Hadoop, Spark

#### LANGUAGES

English/Russian bilingual

#### PROFESSIONAL SERVICES AND AFFILIATIONS

Program Committee: SIGIR 2007-2008, IEEE BigData 2015-2016, KDD 2016

Conference Review: IEEE ISI 2005, InfoScale 2007, ICDM (2005,2010,2011)

Journal Review: Information Retrieval 2007,TKDE (2010-), KAIS (2013-), DAMI (2014-), TNN (2015-)

Member of ACM SIGKDD

#### COMPETITION RESULTS

- Member of winning team at "e-LICO multi-omics prediction challenge with background knowledge on Obstructive Nephropathy" (<http://tunedit.org/challenge/ON>), 2010
- Runner-up on Pulmonary Embolism (PE) and Patient Classification tasks of KDD Cup 2006.
- Member of MMS DIMACS team in Simulated Entity Resolution KDD Challenge 2005 - best and runner-up results on majority of the tasks.
- Member of top scoring team (MMS DIMACS) in TREC'04 Genomics Track.

#### PATENTS GRANTED

- Dmitriy Fradkin, Fabian Moerchen. Condition monitoring with automatically generated error templates from log messages and sensor trends based on time semi-intervals. US 8423493.
- Razvan Ioan Ionasec, Puneet Sharma, Bogdan Georgescu, Andrey Torzhkov, Fabian Moerchen, Gayle M. Wittenberg, Dmitriy Fradkin, Dorin Comaniciu. Multi-Component Heart and Aorta Modelling from High-Resolution MR and CT for Decision Support in Cardiac Disease. US 8527251.
- Dmitriy Fradkin, Fabian Moerchen. Temporal pattern matching in large collections of log messages. US 9026550

#### COLLECTIONS/BOOK CHAPTERS

1. Dmitriy Fradkin, Ilya Muchnik. Support Vector Machines. Eds. J. Abello and G. Carmode. "Discrete Methods in Epidemiology", DIMACS Series in Discrete Mathematics and Theoretical Computer Science, volume 70, pp. 13-20, 2006.

2. Dmitriy Fradkin, Ilya Muchnik, Patrick Hermans, Kenton Morgan. Validation of Epidemiological Models: Chicken Epidemiology in the UK. Eds. J. Abello and G. Carmode. "Discrete Methods in Epidemiology", DIMACS Series in Discrete Mathematics and Theoretical Computer Science, volume 70, pp. 243-256, 2006.

#### JOURNAL PUBLICATIONS

3. Iyad Batal, Gregory Cooper, Dmitriy Fradkin, James Harrison, Fabian Moerchen, Milos Hauskrecht. An Efficient Pattern Mining Approach for Event Detection in Multivariate Temporal Data. Knowledge and Information Systems (KAIS), 2016.
4. Dmitriy Fradkin, Fabian Moerchen. Mining Sequential Patterns for Classification. Knowledge and Information Systems (KAIS), 2015. doi: 10.1007/s10115-014-0817-0
5. Jan Haas et. al. Atlas of the clinical genetics of human dilated cardiomyopathy. European Heart Journal. 2014. <https://doi.org/10.1093/eurheartj/ehu301>
6. Hoang Thanh Lam, Fabian Moerchen, Dmitriy Fradkin, Toon Calders. Mining Compressing Sequential Patterns. Statistical Analysis and Data Mining Journal. 2013. doi: 10.1002/sam.11192
7. Hassan H. Malik, Dmitriy Fradkin, Fabian Moerchen. Single Pass Text Classification by Direct Feature Weighting. Knowledge and Information Systems (KAIS), 2011.
8. Hassan H. Malik, John R. Kender, Dmitriy Fradkin, Fabian Moerchen. Hierarchical Document Clustering using Local Patterns. Data Mining and Knowledge Discovery (DMKD), Volume 21 (1), July 2010. doi: 10.1007/s10618-010-0172-z
9. Patrick G. Hermans, Dmitriy Fradkin, Ilya B. Muchnik, and Kenton L. Morgan Prevalence of wet litter and the associated risk factors in broiler flocks in the United Kingdom. Veterinary Record, May 2006; 158: 615 - 622.

#### REFEREED CONFERENCE AND WORKSHOP PUBLICATIONS

10. Incorporating Task Analysis in the Design of a Tool for a Complex and Exploratory Search Task Tugba Kulahcioglu, Dmitriy Fradkin, Sridharan Palanivelu ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR), March 2017
11. Ruben Sipos, Dmitriy Fradkin, Fabian Moerchen, Zhuang Wang. Log-based Predictive Maintenance. International Conference on Knowledge Discovery and Data Mining (KDD), August 2014.
12. Hoang Thanh Lam, Toon Calders, Jie Yang, Fabian Moerchen and Dmitriy Fradkin. Zips: Mining Compressing Sequential Patterns in Streams. Interactive Data Exploration and Analytics (IDEA) workshop at International Conference on Knowledge Discovery and Data Mining (KDD), August 2013.
13. Iyad Batal, Dmitriy Fradkin, James Harrison, Fabian Moerchen, Milos Hauskrecht. An Efficient Approach for Mining Recent Temporal Patterns. International Conference on Knowledge Discovery and Data Mining (KDD), August 2012.
14. Hoang Thanh Lam, Fabian Moerchen, Dmitriy Fradkin, Toon Calders. Mining Compressing Sequential Patterns. SIAM Conference on Data Mining (SDM), April 2012.
15. Dmitriy Fradkin, Fabian Moerchen. Margin-closed frequent sequential pattern mining Useful Patterns workshop (UP'2010), at International Conference on Knowledge Discovery and Data Mining (KDD), July 2010.
16. Fabian Moerchen, Dmitriy Fradkin. Robust mining of time intervals with semi-interval partial order patterns. SIAM Conference on Data Mining (SDM), April 2010.
17. Fabian Moerchen, Dmitriy Fradkin, Mathaeus Dejori, and Bernd Wachmann. Anticipating Emerging Trends in Biomedical Literature. Annual Symposium on Medical Informatics (AMIA), November 2008.

18. Saikat Mukherjee, Dmitriy Fradkin, Michael Roth. Classifying Spend Transactions with Off-the-Shelf Learning Components. International Conference on Tools with Artificial Intelligence (ICTAI). November 2008.
19. Dmitriy Fradkin. Clustering Inside Classes Improves Performance of Linear Classifiers. International Conference on Tools with Artificial Intelligence (ICTAI). November 2008.
20. Fabian Moerchen, Mathaeus DeJori, Dmitriy Fradkin, Julien Etienne, Bernd Wachmann and Marcus Bundschuh. Anticipating Annotations and Emerging Trends in Biomedical Literature. International Conference on Knowledge Discovery and Data Mining (KDD), August 2008.
21. David Madigan, Alexander Genkin, David D. Lewis and Dmitriy Fradkin. Bayesian Multinomial Logistic Regression for Author identification. MaxEnt 2005.
22. David Madigan, Alexander Genkin, David D. Lewis, Shlomo Argamon, Dmitriy Fradkin and Li Ye. Author Identification on the Large Scale. CSNA 2005.
23. Dmitriy Fradkin and Paul Kantor. Methods for Learning Classifier Combinations: No Clear Winner. ACM Symposium on Applied Computing (SAC), March 2005.
24. Sundara Venkataraman, Dimitris Metaxas, Dmitriy Fradkin, Casimir Kulikowski, Ilya Muchnik. Distinguishing Mislabeled Data from Correctly Labeled Data in Classifier Design. International Conference on Tools with Artificial Intelligence (ICTAI). November 2004.
25. Dmitriy Fradkin and Paul Kantor. A Design Space Approach to Analysis of Information Retrieval Adaptive Filtering Systems. Conference on Information and Knowledge Management (CIKM). November 2004.
26. Dmitriy Fradkin and Ilya B. Muchnik. Clusters With Core-Tail Hierarchical Structure And Their Applications To Machine Learning Classification. International Conference on Integration of Knowledge Intensive Multi-Agent Systems (KIMAS), October 2003.
27. Dmitriy Fradkin, Ilya B. Muchnik and Simon Streltsov. Image Compression in Real-Time Multi-processor Systems Using Divisive K-Means Clustering. International Conference on Integration of Knowledge Intensive Multi-Agent Systems (KIMAS), October 2003.
28. Dmitriy Fradkin and David Madigan. Experiments with Random Projections for Machine Learning. International Conference on Knowledge Discovery and Data Mining (KDD), August 2003.

#### OTHER PUBLICATIONS

29. Andrei Anghelescu, Aynur Dayanik, Dmitriy Fradkin, Alex Genkin, Paul Kantor, David Lewis, David Madigan, Ilya Muchnik, Fred Roberts. Simulated Entity Resolution by Diverse Means: DIMACS Work on the KDD Challenge of 2005. DIMACS Technical Report # 2005-42, December 2005.
30. Dmitriy Fradkin, Dona Schneider and Ilya Muchnik Machine Learning Methods in the Analysis of Lung Cancer Survival Data DIMACS Technical Report # 2005-35, October 2005.
31. Dmitriy Fradkin and Michael Littman. Exploration Approaches to Adaptive Filtering. DIMACS Technical Report # 2005-01, January 2005.
32. Aynur Dayanik, Dmitriy Fradkin, Alex Genkin, Paul Kantor, David D. Lewis, David Madigan, Vladimir Menkov. DIMACS at the TREC 2004 Genomics Track. Proceedings of the 13th Text REtrieval Conference (TREC 2004).
33. Dmitriy Fradkin and Ilya B. Muchnik. A Study of K-Means Clustering for Improving Classification Accuracy of Multi-Class SVM. DIMACS Technical Report # 2004-02, February 2004.