

CHARLES ELKAN

Department of Computer Science and Engineering
University of California, San Diego
La Jolla, CA 92093-0404
(619) 379-9852
elkan@ucsd.edu
January 13, 2012

ACADEMIC EMPLOYMENT

Professor, Department of Computer Science and Engineering, University of California, San Diego, since 2004. Assistant then Associate Professor, 1990 to 2004.

Visiting Scientist, Computer Science and Artificial Intelligence Laboratory, MIT, August 2005 to July 2006.

Visiting Associate Professor, Department of Computer Science, Harvard University, 1998/1999.

Postdoctoral Fellow, Knowledge Representation and Reasoning Group, Department of Computer Science, University of Toronto, 1989/1990.

Lecturer, University of Pennsylvania, 1989 and 1990. Co-taught courses sponsored by the US Army Research Office.

Lecturer, Cornell University, 1987. Designed and co-taught an introductory computer science course.

EDUCATION

Ph.D. in computer science, Cornell University, conferred August 1990. Thesis title *Flexible Concurrency Control by Reasoning about Queries and Updates*. Committee members Prakash Panangaden, Dexter Kozen and George Staller (Department of Economics).

M.S. in computer science, Cornell University, January 1988.

Visiting scholar in artificial intelligence and economics, Stanford University, 1986.

B.A. (Honors) in mathematics, Fitzwilliam College, Cambridge University, June 1984. Focus on optimization and statistics.

Double *maturité* with distinction in natural sciences and distinction in Latin, Collège Rousseau, Geneva, Switzerland, June 1981.

SELECTED HONORS

Keynote speaker, 16th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Kuala Lumpur, Malaysia, May/June 2012.

Invited speaker, *The analytics landscape: A personal view*, Indo-US Workshop on Large Scale Data Analytics and Intelligent Services, Bangalore, India, December 2011.

Keynote speaker, *A vision for reinforcement learning for predictive maintenance*, Workshop on Data Mining for Service and Maintenance, ACM International Conference on Knowledge Discovery in Databases (KDD), August 2011.

Top submission selected for immediate journal publication, *Predicting labels for dyadic data* with Aditya Krishna Menon, European Conference on Machine Learning (ECML), September 2010.

Invited speaker, *Recommender systems*, National Institute of Statistical Science (NISS) Workshop on Computational Advertising, November 2009.

Keynote speaker, *Factorizing matrices with missing entries: Alternative approaches*, Workshop on Data Mining with Matrices and Tensors (DMMT), ACM International Conference on Knowledge Discovery in Databases (KDD), July 2009.

Keynote speaker, *Clustering with k-means: faster, smarter, cheaper*, Workshop on Clustering High-Dimensional Data, SIAM International Conference on Data Mining (ICDM), 2004.

Keynote speaker, *What are the real challenges in data mining?*, Workshop on Learning from Imbalanced Datasets, International Conference on Machine Learning (ICML), August 2003.

Award for first place out of 43 entries in the CoIL Challenge data mining competition, May 2000.

Award for first place out of 45 entries in the data mining competition at the ACM International Conference on Knowledge Discovery in Databases (KDD), August 1997.

Best paper award, IEEE Conference on Artificial Intelligence for Applications (CAIA), March 1993, for *Categorization-based diagnostic problem solving in the VLSI design domain* with Amir Hekmatpour.

Award, best-written paper competition, National Conference on Artificial Intelligence (AAAI), July 1993, for *The paradoxical success of fuzzy logic* (sole author).

RECENT RESEARCH GRANTS

The Transporter Classification Database (TCDB), NIH R01 GM077402-05A1, co-PI with Prof. Milton Saier, Division of Biology, UCSD, \$860,000, 3/1/2011 to 2/28/2015.

Learning from presence-only data, with applications in ecoinformatics and cyber security, University of California/National Laboratory Research Grant, sole PI, \$472,000, 1/1/2009 to 12/31/2011.

Gift to support UCSD data mining contests, Fair Isaac, Inc., sole PI, \$20,000 to \$25,000 per year, 2004, 2005, 2006, 2007, 2008, 2009, 2010.

The Transporter Classification Database (TCDB), NIH R01 GM077402-03, co-PI with Prof. Milton Saier, Division of Biology, UCSD, \$760,000, 4/1/2006 to 3/31/2010.

Learning correct policies despite bias in available training data, California MICRO and Fair Isaac, Inc., sole PI, \$65,000 per year approximately, 2003, 2004, 2005, 2006.

Emergence and development of cognition, Matsushita Electric Industrial Corporation, Osaka, Japan,

co-PI with Prof. Garrison Cottrell, CSE, UCSD, \$195,000, 2001 to 2004.

Data mining for reliability and availability, Sun Microsystems, Inc., \$50,000, 2003 to 2004.

CONSULTING EMPLOYMENT

Member, Advisory Board, Project on Responsible Business and Social Media, Copenhagen Business School, Denmark.

Consultant on real-time video recommendation system design, CullTV Inc., 2010 and 2011.

Consultant on text mining, Qualcomm, Inc., 2010.

Intellectual property litigation expert, Latham & Watkins LLP, 2009.

Consultant on data mining contest design and judging, Netflix Inc., 2006 to 2009.

Consultant on data mining, Intuit Inc., San Diego, 2006 and 2007.

Consultant on research management, Systems Development Laboratory, Hitachi Ltd., Tokyo, 2003 and 2004.

Consultant on machine learning research, Matsushita Electric Industrial Co., Ltd. (now named Panasonic Corporation), Osaka, 2003 and 2004.

Consultant on research management, Hewlett-Packard Laboratories, Palo Alto, 1999 and 2000.

Senior Scientist, Knowledge Stream Partners, Inc., Boston, 1998 and 1999.

Consultant on data mining, Science Applications International Corporation (SAIC), San Diego, 1997.

Consultant on information system design, Sony (San Diego), IBM (San Jose), and Alcoa (San Diego), 1995.

Consultant on software intellectual property issues, Appx Inc. (Richmond, Virginia) and Morris, Manning & Martin, Attorneys at Law (Atlanta, Georgia), 1995.

Research intern, Intelligent Systems Laboratory, Xerox Palo Alto Research Center, 1988.

Member, technical staff, Statistics Department, Mathematics Research Center, ATT Bell Laboratories, Murray Hill, New Jersey, 1986.

Consultant on quality control, Katalco Corporation, Chicago, 1983.

Applications programmer, CompuSoft S.A., Geneva, Switzerland, 1982.

RECENT UNIVERSITY SERVICE

Chair, CSE M.S. committee, 2000 to 2011.

Chair, UCSD Academic Administrator Review Panel, 2007 to present.

Member, Academic Senate Committee on Budget Transparency, 2010 to present.

Chair, Review Committee for the Provost of Warren College, 2009.

Chair, Academic Senate Library Committee, 2001/02 and 2007/08. Vice chair, 2000/01. Member, 1999/2000.

Member, Academic Senate Committee on Educational Policy, 2004/05.

Member, Academic Senate Council, 2003/2004.

Chair, Academic Senate Committee on Affirmative Action and Diversity, 2003/04. Member, 2002/03.

Organizer, interdisciplinary Ph.D. program in cognitive science seminar series on *Historical and Conceptual Foundations of Cognitive Science*, fall 2001.

RECENT PROFESSIONAL SERVICE

Area program chair, International Conference on Machine Learning (ICML), 2011 and 2012.

Area program chair, ACM International Conference on Knowledge Discovery in Databases (KDD), 2011 and 2012.

Current action editor, *ACM Transactions on Intelligent Systems and Technology* and *Journal of Machine Learning Research*.

Former action editor or editorial board member, *Computational Intelligence*, *Journal of Data Mining and Knowledge Discovery*, *Machine Learning*, *Journal of Artificial Intelligence Research*, and *IEEE Transactions on Data and Knowledge Engineering*.

Previously area chair or program committee member for major conferences in computer science over 40 times from 1990 to 2011.

Reviewer of papers submitted to dozens of journals including *ACM Transactions on Database Systems*, *Annals of Pure and Applied Logic*, *Computer Applications in the Biological Sciences*, *Pattern Recognition*, and *Quantitative Finance*.

Referee for appointments or promotions as associate or full professor at Columbia University, New York University, Northwestern University, and numerous other universities.

PH.D. GRADUATES

Amir Hekmatpour, Ph.D. conferred June 1993. Thesis title *A methodology and architecture for interactive knowledge-based diagnostic problem-solving in VLSI manufacturing*. Recently senior engineer, IBM Microelectronics. Previously senior lecturer, Department of Electrical and Computer Engineering, University of Texas, Austin.

Timothy L. Bailey, Ph.D. conferred June 1995. Thesis title *Discovering motifs in DNA and protein sequences: The approximate common substring problem*. Now senior research fellow, University of Queensland, Australia.

Michael Sussna, Ph.D. conferred June 1997. Thesis title *Text retrieval using inference in semantic metanetworks*.

Alvaro Monge, Ph.D. conferred August 1997. Thesis title *Adaptive detection of approximately du-*

plicate database records and the database integration approach to information discovery. Now full professor, Department of Computer Engineering and Computer Science, California State University, Long Beach.

William Stafford Noble, Ph.D. conferred June 1998. Thesis title *A Bayesian approach to motif-based protein modeling*. Now full professor, Departments of Genome Sciences and of Computer Science, University of Washington, Seattle.

Greg Hamerly, Ph.D. conferred June 2003. Thesis title *Learning structure and concepts in data through data clustering*. Now associate professor with tenure, Department of Computer Science, Baylor University.

Bianca Zadrozny, Ph.D. conferred August 2003. Thesis title *Policy mining: Learning decision policies from fixed sets of data*. Now permanent research staff member, IBM Research, Rio de Janeiro, Brazil. Previously associate professor with tenure, Computer Science Department, Fluminense Federal University, Brazil.

David Kauchak, Ph.D. conferred June 2006. Thesis title *Contributions to research on machine translation*. Now assistant professor, Middlebury College. Previously visiting assistant professor, Pomona College.

Eric Wiewiora, Ph.D. conferred December 2007. Thesis title *Modeling probability distributions with predictive state representations*. Now engineering scientist at Intellisis Inc., San Diego.

Doug Turnbull, Ph.D. conferred July 2008. Thesis title *Design and development of a semantic music discovery engine*. Now assistant professor, Ithaca College. Previously visiting assistant professor, Swarthmore College.

PEER-REVIEWED PUBLICATIONS

Charles Elkan, David Lubinsky, and Daryl Pregibon. Automated descriptions of data. In *Proceedings of the Fifth International Symposium on Data Analysis and Computer Science*, pp. 165–170. Versailles, France, September 1987.

Charles Elkan and David McAllester. Automated inductive reasoning about logic programs. In *Proceedings of the Fifth International Conference Symposium on Logic Programming*, pp. 876–892. Seattle, August 1988. MIT Press.

Charles Elkan. A decision procedure for conjunctive query disjointness. In *Proceedings of the Eighth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS)*, pp. 134–139. Philadelphia, March 1989. ACM Press.

Charles Elkan. Conspiracy numbers and caching for searching and/or trees and theorem-proving. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 341–346. Detroit, August 1989. Morgan Kaufmann Publishers, Inc.

Charles Elkan. Logical characterizations of nonmonotonic TMSs. In *Proceedings of the Symposium on Mathematical Foundations of Computer Science*, pp. 218–224. Pořąbka-Kozubnik, Poland,

August 1989. Springer Verlag Lecture Notes in Computer Science, no. 379.

Charles Elkan. Independence of logic database updates and queries. In *Proceedings of the Ninth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS)*, pp. 154–160. Nashville, Tennessee, April 1990. ACM Press.

Charles Elkan. A rational reconstruction of nonmonotonic truth maintenance systems. *Artificial Intelligence*, **43(2)**219–234, May 1990.

Charles Elkan. Incremental, approximate planning. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pp. 145–150. Boston, August 1990. MIT Press.

Alberto Segre, Charles Elkan, and Alex Russell. A critical look at experimental evaluations of EBL. *Machine Learning*, **6(2)**183–195, February 1991.

Charles Elkan. Formalizing causation in first-order logic: Lessons from an example. Published at the *AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning*, pp. 41–47. Stanford, California, March 1991.

Russell Greiner and Charles Elkan. Measuring and improving the effectiveness of representations. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 518–524. Sydney, Australia, August 1990. Morgan Kaufmann Publishers, Inc.

Alberto M. Segre, Charles Elkan, Daniel Scharstein, Geoffrey J. Gordon, and Alexander Russell. Adaptive inference. In *Foundations of Knowledge Acquisition, Vol. II: Machine Learning: Induction, Analogy, and Discovery*, edited by Susan Chipman and Alan Meyrowitz, pp. 43–81. Kluwer Academic Publishers, 1992.

Charles Elkan. Reasoning about action in first-order logic. In *Proceedings of the Ninth Biennial Conference of the Canadian Society for Computational Studies of Intelligence (CSCSI)*, pp. 221–227. Vancouver, Canada, May 1992. Morgan Kaufmann Publishers, Inc.

Amir Hekmatpour and Charles Elkan. Categorization-based diagnostic problem solving in the VLSI design domain. In *Proceedings of the Ninth IEEE Conference on Artificial Intelligence for Applications (CAIA)*, pp. 121–127. Orlando, Florida, March 1993. IEEE Press. Winner, best paper award.

Amir Hekmatpour and Charles Elkan. A multimedia expert system for wafer polisher maintenance. In *Proceedings of the Ninth IEEE Conference on Artificial Intelligence for Applications (CAIA)*, p. 453. Orlando, Florida, March 1993. IEEE Press.

Charles Elkan. The paradoxical success of fuzzy logic. In *Proceedings of the Eleventh National Conference on Artificial Intelligence (AAAI)*, pp. 698–703. Boston, Massachusetts, July 1993. MIT Press. Award for best-written paper.

Timothy L. Bailey and Charles Elkan. Estimating the accuracy of learned concepts. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 895–900. Chambéry, France, September 1993. Morgan Kaufmann Publishers, Inc.

Alberto M. Segre and Charles Elkan. A high-performance explanation-based learning algorithm, *Artificial Intelligence*, **69(1)**1–50, September 1994.

Timothy L. Bailey and Charles Elkan. Cross-validation and modal theories. In *Computational Learning Theory and Natural Learning Systems, Volume III: Selecting Good Models*, ed. Thomas Petsche, Stephen J. Hanson, and Jude W. Shavlik, pp. 345–359. MIT Press, 1995.

Charles Elkan. The paradoxical success of fuzzy logic. *IEEE Expert*, **6(6)**3–8, August 1994. With fifteen responses by other researchers on pp. 9–46.

Charles Elkan. The paradoxical controversy over fuzzy logic. *IEEE Expert*, **6(6)**47–49, August 1994.

Timothy L. Bailey and Charles Elkan. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology (ISMB)*, pp. 28–36. Stanford, California, August 1994. AAAI Press.

Timothy L. Bailey and Charles Elkan. Unsupervised learning of multiple motifs in biopolymers using expectation maximization. *Machine Learning*, **21(1–2)**51–80, October 1995.

Timothy L. Bailey and Charles Elkan. The value of prior knowledge in discovering motifs with MEME. In *Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology (ISMB)*, pp. 21–29. Cambridge, England, July 1995. AAAI Press.

Alberto M. Segre, Geoffrey J. Gordon, and Charles Elkan. Exploratory analysis of speedup learning data using expectation maximization. *Artificial Intelligence*, **85(1-2)**301–319, August 1996.

Charles Elkan. Reasoning about unknown, counterfactual, and nondeterministic actions in first-order logic. In *Proceedings of the Eleventh Biennial Conference of the Canadian Society for Computational Studies of Intelligence (CSCSI)*, pp. 54–68. Toronto, Canada, May 1996. Springer Verlag.

Karan Bhatia and Charles Elkan. LPMEME: A statistical method for inductive logic programming. In *Proceedings of the Eleventh Biennial Conference of the Canadian Society for Computational Studies of Intelligence (CSCSI)*, pp. 227–239. Toronto, Canada, May 1996. Springer Verlag.

William N. Grundy, Timothy L. Bailey, and Charles Elkan. ParaMEME: a parallel implementation and a web interface for a DNA and protein motif discovery tool. *Computer Applications in the Biosciences*, **12(4)**303–310, 1996.

Alvaro E. Monge and Charles Elkan. The field matching problem: Algorithms and applications. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 267–270. Portland, Oregon, August 1996. AAAI Press.

Timothy L. Bailey, Michael E. Baker, and Charles Elkan. An artificial intelligence approach to motif discovery in protein sequences: Application to steroid dehydrogenases. *Journal of Steroid Biochemistry and Molecular Biology*, **62(1)**:29–44, 1997.

William N. Grundy, Timothy L. Bailey, Charles Elkan, and Michael E. Baker. Hidden Markov

model analysis of motifs in steroid dehydrogenases and their homologs. *Biochemical and Biophysical Research Communications*, **231(3)**:760–766, 1997.

William N. Grundy, Timothy L. Bailey, Charles Elkan, and Michael E. Baker. Meta-MEME: Motif-based hidden Markov models of protein families. *Computer Applications in the Biosciences*, **13(4)**:397–406, 1997.

Alvaro E. Monge and Charles Elkan. An efficient domain-independent algorithm for detecting approximately duplicate database records. Published at the *SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD)*. Tucson, Arizona, May 1997.

Timothy L. Bailey, Michael E. Baker, Charles Elkan, and William N. Grundy. MEME, MAST, and Meta-MEME: New tools for motif discovery in protein sequences. In *Pattern Discovery in Biomolecular Data: Tools, Techniques, and Applications*, J. Wang, B. Shapiro and D. Shasha, editors. Oxford University Press, 1999.

Michael E. Baker, William N. Grundy, and Charles Elkan. Spinach CSP41, an mRNA-binding protein and ribonuclease, is homologous to nucleotide-sugar epimerases and hydroxysteroid dehydrogenases. *Biochemical and Biophysical Research Communications*, **248(2)**:250–254, 1998.

Michael E. Baker, William N. Grundy, and Charles Elkan. A common ancestor for a subunit in the mitochondrial proton-translocating NADH:ubiquinone oxidoreductase (complex I) and short-chain dehydrogenases/reductases. *Cellular and Molecular Life Sciences*, **55(3)**:450–455, 1999.

Fredrik Farnstrom, James Lewis, and Charles Elkan. Scalability for clustering algorithms revisited. *ACM SIGKDD Explorations*, **2(1)**:51–57, August 2000.

Charles Elkan. Paradoxes of fuzzy logic, revisited. *International Journal of Approximate Reasoning*, **26(2)**:153–155, 2001.

Koji Morikawa, Sameer Agarwal, Charles Elkan, and Garrison W. Cottrell. A taxonomy of computational and social learning. In *Proceedings of the Workshop on Developmental Embodied Cognition (DECO)*, Edinburgh, July 2001, pp. 46–50.

Greg Hamerly and Charles Elkan. Bayesian approaches to failure prediction for disk drives. In *Proceedings of the 18th International Conference on Machine Learning (ICML)*, July 2001, pp. 202–209.

Bianca Zadrozny and Charles Elkan. Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers. In *Proceedings of the 18th International Conference on Machine Learning (ICML)*, July 2001, pp. 609–616.

Charles Elkan. The foundations of cost-sensitive learning. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI)*, August 2001, pp. 973–978.

Bianca Zadrozny and Charles Elkan. Learning and making decisions when costs and probabilities are both unknown. In *Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining (KDD)*, September 2001, pp. 204–213.

Charles Elkan. Magical thinking in data mining: Lessons from CoIL Challenge 2000. In *Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining (KDD)*, September 2001, pp. 426–431.

Bianca Zadrozny and Charles Elkan. Transforming classifier scores into accurate multiclass probability estimates. In *Proceedings of the Eighth International Conference on Knowledge Discovery and Data Mining (KDD)*, August 2002, pp. 694–699.

Gordon F. Hughes, Joseph F. Murray, Kenneth Kreutz-Delgado, and Charles Elkan. Improved disk-drive failure warnings. *IEEE Transactions on Reliability*, **51(3)**:350–357, September 2002.

Greg Hamerly and Charles Elkan. Alternatives to the k -means algorithm that find better clusterings. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM)*, November 2002, pp. 600–607.

Charles Elkan. Using the triangle inequality to accelerate k -means. In *Proceedings of the 20th International Conference on Machine Learning (ICML)*, Washington, DC, August 2003, pp. 147–153.

Eric Wiewiora, Garrison Cottrell, and Charles Elkan. Principled methods for advising reinforcement learning agents. In *Proceedings of the 20th International Conference on Machine Learning (ICML)*, Washington, DC, August 2003, pp. 792–799.

David Kauchak and Charles Elkan. Learning rules to improve a machine translation system. In *Proceedings of the 14th European Conference on Machine Learning (ECML)*, Dubrovnik, Croatia, September 2003, pp. 205–216.

Greg Hamerly and Charles Elkan. Learning the k in k -means. In *Proceedings of the 17th Annual Conference on Neural Information Processing Systems (NIPS)*, December 2003.

Andrew Smith and Charles Elkan. A Bayesian network framework for reject inference. In *Proceedings of the Tenth International Conference on Knowledge Discovery and Data Mining (KDD)*, August 2004, pp. 286–295.

David Kauchak, Joseph Smarr, and Charles Elkan. Sources of success for boosted wrapper induction. *Journal of Machine Learning Research*, **5**:499–527, 2004.

Doug Turnbull and Charles Elkan. Fast recognition of musical genre using RBF networks. *IEEE Transactions on Data and Knowledge Engineering*, **17(4)**:580–584, 2005.

Rasmus Madsen, David Kauchak, and Charles Elkan. Modeling word burstiness using the Dirichlet distribution. In *Proceedings of the 22nd International Conference on Machine Learning (ICML)*, Bonn, Germany, August 2005, pp. 545–552.

Charles Elkan. Deriving TF-IDF as a Fisher kernel. In *Proceedings of the 12th International Conference on String Processing and Information Retrieval (SPIRE)*, Buenos Aires, Argentina, November 2–4, 2005, Springer Lectures Notes in Computer Science, vol. 3772, pp. 295–300.

Charles Elkan. Clustering documents with an exponential-family approximation of the Dirichlet

compound multinomial distribution In *Proceedings of the 23rd International Conference on Machine Learning (ICML)*, Pittsburgh, PA, August 2006, pp. 289–296.

Charles Elkan. Method and system for selecting documents by measuring document quality. United States Patent 7200606, issued April 3, 2007.

Andrew Smith and Charles Elkan. Making generative classifiers robust to selection bias. In *Proceedings of the 13th International Conference on Knowledge Discovery and Data Mining (KDD)*, August 2007, pp. 657–666.

Sanmay Das, Milton H. Saier, and Charles Elkan. Finding transport proteins in a general protein database. In *Proceedings of the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)*, Warsaw, Poland, September 2007, pp. 54–66.

Charles Elkan and Keith Noto. Learning classifiers from only positive and unlabeled data. In *Proceedings of the 14th International Conference on Knowledge Discovery and Data Mining (KDD)*, Las Vegas, August 2008, pp. 213–220.

Guilherme Hoefel and Charles Elkan. Learning a two-stage SVM/CRF sequence classifier. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM)*, Napa, October 2008, pp. 271–278.

Keith Noto, Milton H. Saier Jr., and Charles Elkan. Learning to find relevant biological articles without negative training examples. In *Proceedings of the 21st Australasian Joint Conference on Artificial Intelligence (AI)*, Auckland, New Zealand, December 2008, pp. 202–213.

Milton H. Saier, Jr., Ming-Ren Yen, Keith Noto, Dorjee Tamang, and Charles Elkan. The transporter classification database: Recent advances. *Nucleic Acids Research*, **37**:274–278, 2008, doi:10.1093/nar/gkn862.

Gabriel Doyle and Charles Elkan. Accounting for burstiness in topic models. In *Proceedings of the 26th International Conference on Machine Learning (ICML)*, June 2009, pp. 36–43.

Gabriel Doyle and Charles Elkan. Financial topic models. *Workshop on Applications for Topic Models: Text and Beyond*, NIPS, December 2009.

Irene Rodriguez-Lujan, Ramon Huerta, Charles Elkan, and Carlos Santa Cruz. Quadratic programming feature selection. *Journal of Machine Learning Research*, **11**:1491–1516, 2010.

Nikos Trovkanis and Charles Elkan. Conditional random fields for word hyphenation. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*, July 2010, pp. 366–374.

Aditya Krishna Menon and Charles Elkan. Predicting labels for dyadic data. *Journal of Data Mining and Knowledge Discovery*, **21**(2):327–343, 2010.

Padhraic Smyth and Charles Elkan. Creativity helps influence prediction precision. *Communications of the ACM*, **53**(4):88, 2010.

Luigi Cerulo, Charles Elkan, and Michele Ceccarelli. Learning gene regulatory networks from only

positive and unlabeled data. *BMC Bioinformatics*, **11**:228, 2010.

Aditya Krishna Menon and Charles Elkan. A log-linear model with latent features for dyadic prediction. In *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, Sydney, Australia, December 2010.

Avinash Atreya and Charles Elkan. Latent semantic indexing (LSI) fails for TREC collections. *ACM SIGKDD Explorations*, **12(2)**:5–10, December 2010.

Wenkai Li, Qinghua Guo, and Charles Elkan. A positive and unlabeled learning algorithm for one-class classification of remote-sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, **49(2)**:717–725, 2011.

Charles Elkan. Preserving privacy in data mining via importance weighting. In *Proceedings of the ECML/PKDD Workshop on Privacy and Security Issues in Data Mining and Machine Learning (PSDML)*, Lecture Notes on AI 6549, pp. 15–21, Springer Verlag, 2011.

Aditya Krishna Menon and Charles Elkan. Fast algorithms for approximating the singular value decomposition. *ACM Transactions on Knowledge Discovery from Data*, **5(2)**:1–36, February 2011.

Aditya Kumar Sehgal, Sanmay Das, Keith Noto, Milton Saier, and Charles Elkan. Identifying relevant data for a biological database: Handcrafted rules versus machine learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **8(3)**:851–857, May-June 2011.

Wenkai Li, Qinghua Guo, and Charles Elkan. Can we model the probability of presence of species without absence data? *Ecography*, published online May 2011. doi:10.1111/j.1600-0587.2011.06888.x.

Martin Krallinger, Miguel Vazquez, Florian Leitner, David Salgado, Andrew Chatranyamontri, Andrew Winter, Livia Perfetto, Leonardo Briganti, Luana Licata, Marta Iannuccelli, Luisa Castagnoli, Gianni Cesareni, Mike Tyers, Gerold Schneider, Fabio Rinaldi, Robert Leaman, Graciela Gonzalez, Sergio Matos, Sun Kim, W. John Wilbur, Luis Rocha, Ashish V Tendulkar, Aniket Rangrej, Vishal Raut, Shashank Agarwal, Feifan Liu, Xinglong Wang, Rafal Rak, Keith Noto, **Charles Elkan**, Zhiyong Lu, Rezarta Islamaj Dogan, Jean-Fred Fontaine, Miguel A. Andrade-Navarro and Alfonso Valencia. The Protein-Protein Interaction Tasks of BioCreative III: Classification/Ranking of Articles and Linking Bio-ontology Concepts to Full Text. Accepted by *BMC Bioinformatics*, to appear in 2011.

Aditya Krishna Menon and Charles Elkan. Link prediction via matrix factorization. In *Proceedings of the European Conference on Machine Learning (ECML)*, September 2011, pp. 1–16.