

David Cieslak

384 Fitzpatrick Hall - Notre Dame, IN 46556

Mobile Phone: 630-965-4686

e-mail: dcieslak@nd.edu

Education

Ph.D. in Computer Science and Engineering, University of Notre Dame, South Bend, Indiana, June 2009, *Dissertation: "Finding Problems in, Proposing Solutions to, and Performing Analysis on Imbalanced Data"*

M.S. in Computer Science and Engineering, University of Notre Dame, South Bend, Indiana, May 2006, *Thesis: "A Clustering Defense Against Distributed Denial of Service Attacks"*

B.S. in Computer Science and Engineering, University of Notre Dame, South Bend, Indiana, May 2004, *GPA: 3.781*

Research Topics

Class Imbalanced Learning
Localized Sampling Optimization
Loss Analysis & Evaluation Metrics and Methods
Cost-Sensitive Learning Methods
Learning under Non-stationary Distributions
Sample Selection Bias
Applications of Above: Finance, Medicine, Systems

Awards

2007-2008 Graduate Research Excellence Award
2006-2007 Outstanding Graduate Teaching Assistant
ICDM 2007 Student Travel Grant
Arthur J. Schmitt Fellowship Recipient

Work and Research Experience

University of Notre Dame, Notre Dame, IN (Summer 2009-present)
Position: Research Scientist, Interdisciplinary Center for Network Science and Applications, Computer Science and Engineering
Responsibilities: Member of corporate collaboration, enabling data mining and analytics for a Fortune 500 and various other companies.. Fused campus research and cyber-infrastructure. Independent data mining and machine learning researcher. Course instructor.

Sandia National Laboratories, Livermore, CA (Summer 2008)

Position: Post-Masters Intern, Informatics and Decision Sciences

Responsibilities: Performed research on the class imbalance problem. Developed a solution combining my research with my mentor's. Implemented method within lab's software tool kit.

University of Notre Dame, Notre Dame, IN (Fall 2004-Summer 2009)

Position: Graduate Research Assistant, Computer Science

Responsibilities: Pursued research within the domains of data mining and machine learning. Published results in several competitive journals and conferences relevant to the field.

Motorola, Libertyville, IL (Summer 2004)

Position: Intern, PCS systems engineering group

Developed a predictive model for cell phone memory usage. Created a web-interface for intuitive interpretation and analysis and incorporating data for model training.

GE Medical Systems, Waukesha, WI (Summer 2003)

Position: EID Intern

Responsibilities: Performed a pilot study to assess feasibility in replacing a licensed database with an internally developed system. The survey developed a working prototype, leading to an eventual implementation and generating a substantial cost-savings.

Progeny Software, South Bend, IN (Summer 2002)

Position: Intern

Responsibilities: Incorporated Microsoft ODBC compatibility in company's flagship software product, making it a viable solution for a wider range of clinics. Also installed Oracle on the company's local database server.

Publications

Journal articles

D. A. Cieslak, N. V. Chawla, W. P. Kegelmeyer, Hellinger
Distance Decision Trees are Robust and Skew-Insensitive, *Data Mining and Knowledge Discover*, 2011

D. A. Cieslak, N. V. Chawla, "A Framework for Monitoring Classifiers' Performance: When and Why Failure Occurs?", *Knowledge and Information System*, 18(1), pg. 83-108, 2009

Q. Liao, D. A. Cieslak, A. Striegel, N. V. Chawla, "Using Selective, Short-Term Memory to Improve Resilience versus DDoS Exhaustion Attacks," *Security and Communication Networks*, 1(4), pg. 287-299, 2008

N. V. Chawla, D. A. Cieslak, L. Hall, A. Joshi, "Automatically Countering Imbalance and Its Empirical Relationship to Cost," *Data Mining and Knowledge Discovery*, 17(2), pg. 225-252, 2008

C. Mano, A. Blaich, Q. Liao, Y. Jiang, D. A. Cieslak, D. Salyers, A. Striegel, "RIPPS: Rogue Identifying Packet Payload Slicer Detecting Unauthorized Wireless Hosts Through Network Traffic Conditioning," *ACM Transactions on Information and System Security*, 11(2), pg. 1-23, 2008

Peer-reviewed Conferences

T. Wright, D. A. Cieslak, G. Madey, S. Arifin, "Garki Online: E-Science Cyberinfrastructure for Research Data on Malaria Transmission," International Workshop on Portals for Life Sciences, September 2009

D. A. Cieslak, N. V. Chawla, "Start Globally, Optimize Locally, Predict Globally: Improving Performance on Unbalanced Data," *IEEE International Conference on Data Mining (ICDM)*, Pisa, Italy, December 2008, pg. 143-152

D. A. Cieslak, N. V. Chawla, D. Thain, "Troubleshooting Thousands of Jobs on Production Grids Using Data Mining Techniques," *IEEE/ACM International Conference on Grid Computing (GRID)*, Tsukuba, Japan, September 2008, pg. 217-224

D. A. Cieslak, N. V. Chawla, "Learning Decision Trees for Unbalanced Data," *European Conference on Machine Learning (ECML)*, Antwerp, Belgium, September 2008, pg. 241-256

D. A. Cieslak, N. V. Chawla, "Analyzing Classifier Performance on Imbalanced Datasets when Training and Testing Distributions Differ," *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, Nanjing, China, May, 2008, pg. 519-526

N. V. Chawla, D. Thain, R. Lichtenwalter, D. A. Cieslak, "Troubleshooting the Grid," NSF Next Generation Systems Workshop, 2008

D. A. Cieslak, N. V. Chawla, "Detecting Fractures in Classifier Performance," *IEEE International Conference on Data Mining (ICDM)*, Omaha, October 2007, pg. 123-132

N. V. Chawla, D. A. Cieslak, "Evaluating Calibration of Probability Estimation Trees," *AAAI Workshop on the Evaluation Methods in Machine Learning*, Boston, July 2006, pg. 18-23

D. A. Cieslak, D. Thain, N. V. Chawla, "Troubleshooting Distributed Systems via Data Mining," *Hot Topics Sessions: 15th IEEE International Symposium on High Performance Distributed Computing (HPDC-15)*, Paris, France, June 2006, pg. 309-312

D. A. Cieslak, N. V. Chawla, A. Striegel, "Combating Imbalance in Network Intrusion Datasets," *IEEE International Conference on Granular Computing*, Athens, Georgia, May 2006, pg. 732-737

Invited Talks

"Trouble Shooting Thousands of Jobs on Production Grids Using Data Mining"

Center for Research Computing Workshop on Scientific Computing

University of Notre Dame

"Learning When Data is Non-Stationary"

Invited Seminar

Sandia National Laboratories, Livermore, CA

"Data Mining on the Grid and for the Grid"

Super Computing 2007

"Data Mining IN Condor and Data Mining ON Condor"

Condor Week 2007

University of Wisconsin

Teaching Experience

CSE40647/60647 Data Mining (TA: Spring 2005, Course Instructor: Spring 2010)

Level: Mixed undergraduate and graduate

Description: Advanced course focusing on data mining, the process of automatic discovery of useful information, patterns, associations, and anomalies. The course primarily covers the machine learning aspect of data mining, including references to probability, statistics, pattern recognition, and information as

necessary. Students complete regular homework assignments, exams, and quizzes; present real applications of data mining; and apply course concepts to a topic of their choosing for the final project.

CSE20212 Fundamentals of Computing II (TA: Spring 2006)

Level: Undergraduate

Description: The second part of a two-course introduction to computing sequence, intended primarily for Computer Science and Computer Engineering majors. The class focuses on object-oriented programming and elementary data structure usage and design in C++. In addition to weekly lab assignments and exams, students apply what they have learned in a large-scale project.

Reviewer

- IEEE Transactions on Systems, Man and Cybernetics
- SIAM International Conference on Data Mining
- IEEE Symposium on Computational Intelligence and Data Mining International Joint Conference on Neural Networks
- ACM Conference on Information and Knowledge Management
- International Joint Conference on Statistical and Syntactical and Structural Pattern Recognition (S+SSPR)
- IEEE International Conference on Machine Learning
- Statistical Analysis and Data Mining
- ACM SIGKDD Conference on Knowledge Discovery and Data Mining

**Program
Committee
Member**

- Conference on Information and Knowledge Management (2009)
- Pacific-Asia Conference on Knowledge Discovery and Data Mining (2009) Workshop on Data Mining when Classes are Imbalanced and Errors Have Cost