

REID A. JOHNSON

Dept. of Computer Science and Engineering
University of Notre Dame
32 Fischer Graduate Res., Apt. 2B
Notre Dame, IN 46556

Office: 384L Nieuwland Hall
Mobile: (217) 761-6047
Email: rjohns15@nd.edu
Website: <http://nd.edu/~rjohns15/>

RESEARCH INTERESTS

- Imbalanced data: binary and multiclass imbalanced supervised learning problems.
- Education and learning analytics: student retention modeling; evaluating learning outcomes.
- Computational social science: impact prediction; social network analysis; science of science.
- Big data analytics: high-performance data mining; MapReduce/Hadoop; Apache Spark.

EDUCATION

Ph.D. in Computer Science and Engineering University of Notre Dame <i>Dissertation Title:</i> Data Science for Imbalanced Data: Methods and Applications <i>Advisor:</i> Prof. Nitesh V. Chawla	2010–2016
B.S. in Computer Science, <i>Summa Cum Laude</i> University of Illinois at Springfield	2007–2009
St. Olaf College	2005–2007

PROFESSIONAL EXPERIENCE

Dept. of Computer Science and Engineering | University of Notre Dame **Notre Dame, IN**
Research Assistant 2010–Present

My research focused on building machine learning models to solve complex longitudinal problems with applications to areas as diverse as ecological informatics, healthcare analytics, scientific impact prediction, and predictive learning analytics. Specific contributions include:

- Created enhancements to algorithms for imbalanced data, with focus on two-class problems.
- Employed a novel algorithm for learning species distributions, drawing from techniques for class imbalance to meet and exceed the performance of the state of the art.
- Analyzed academic social networks to identify key publication and collaboration signatures predictive of scientific impact, and used the insights to develop an impact prediction tool.
- Collaborated in the creation and evaluation of a novel multi-relational link prediction algorithm to predict drug-disease interactions, and developed an interactive web-based framework to visualize the predictions generated by the method.

Data Science for Social Good Fellowship | University of Chicago

Chicago, IL

Data Science Fellow

Summer 2015

- Worked as part of a small, interdisciplinary team to predict on-time high school graduation from student data provided by several partnering U.S. public school districts.
- Used PostgreSQL and standard Python tools and libraries to develop a data-driven predictive model to identify academically at-risk students that has the potential to increase identification of at-risk students by up to 80% if implemented.

TEACHING

University of Notre Dame

Notre Dame, IN

Course Instructor

Machine Learning (CSE 40625)

Spring 2017

Data Science (CSE 44648)

Summer 2016

- Instructed the inaugural course on data science, offered online.
- Helped develop the course content and held regular teaching live sessions.

Data Mining (CSE 40647)

Spring 2014

- Co-instructed the annual offering on data mining to undergraduate/graduate students.
- Revised the course content to include hands-on experience with Python programming.
- Developed and delivered lecture material and Python-based demonstrations.

Teaching Assistant

Ethics and Professional Issues (CSE 40175)

Spring 2011

Discrete Mathematics (CSE 20110)

Fall 2010

Volunteer Teaching

The Stanley Clark School

Summer 2014, 2016

- Instructed a week-long course introducing youth to basic programming concepts.

AWARDS AND HONORS

Second Place Team, CNT Urban Sustainability Apps Competition	2015
Best Paper Award Nomination, WSDM 2015	2015
Outstanding Student Service Award	2014
Best Poster Award (Student Choice), 8 th Annual CSE-SRS Poster Symposium	2013
Undergraduate Honors in Computer Science	2010

REFEREED PUBLICATIONS

1. **Reid A. Johnson**, Jason D. K. Dzurisin, Nitesh V. Chawla. Large-Scale Machine Learning for Species Distributions. *Large-Scale Machine Learning in the Earth Sciences*, Submitted.
2. Yuxiao Dong, **Reid A. Johnson**, Nitesh V. Chawla. Can Scientific Impact Be Predicted? *IEEE Transactions on Big Data (TBD)*, 2016.

3. Andrea Dal Pozzolo, Olivier Caelen, **Reid A. Johnson**, Gianluca Bontempi. Calibrating Probability with Undersampling for Unbalanced Classification. In *Proceedings of the IEEE Symposium Series on Computational Intelligence (SSCI)*, 2015.
4. **Reid A. Johnson**, Ruobin Gong, Siobhan Greatorex-Voith, Anushka Anand, Alan Fritzler. A Data-Driven Framework for Identifying High School Students at Risk of Not Graduating On Time. *Bloomberg Data for Good Exchange*, 2015.
5. Yuxiao Dong, **Reid A. Johnson**, Yang Yang, Nitesh V. Chawla. Collaboration Signatures Reveal Scientific Impact. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 480–487, 2015.
6. **Reid A. Johnson**, Troy Raeder, Nitesh V. Chawla. Optimizing Classifiers for Hypothetical Scenarios. *Advances in Knowledge Discovery and Data Mining (PAKDD)*, pp. 264–276, 2015.
7. Yuxiao Dong, **Reid A. Johnson**, and Nitesh V. Chawla. Will This Paper Increase Your *h*-index? Scientific Impact Prediction. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (WSDM)*, pp. 149–158, 2015. **Best Paper Award Nomination**
8. Andrea Dal Pozzolo, **Reid A. Johnson**, Olivier Caelen, Serge Waterschoot, Nitesh V. Chawla, and Gianluca Bontempi. Using HDDT to Avoid Instances Propagation in Unbalanced and Evolving Data Streams. In *Proceedings of the Twenty-Fourth International Joint Conference on Neural Networks (IJCNN)*, pp. 588–594, 2014.
9. Andrew K. Rider, **Reid A. Johnson**, Darcy A. Davis, T. Ryan Hoens, and Nitesh V. Chawla. Classifier Evaluation with Missing Negative Class Labels. In *Proceedings of the Twelfth International Conference on Intelligent Data Analysis (IDA)*, pp. 380–391, 2013.
10. **Reid A. Johnson**, Nitesh V. Chawla, and Jessica J. Hellmann. Species Distribution Modeling and Prediction: A Class Imbalance Problem. In *Proceedings of the 2012 Conference on Intelligent Data Understanding (CIDU)*, pp. 9–16, 2012.
11. **Reid A. Johnson**, Yang Yang, Everaldo Aguiar, Andrew Rider, and Nitesh V. Chawla. ALIVE: A Multi-Relational Link Prediction Environment for the Healthcare Domain. In *Emerging Trends in Knowledge Discovery and Data Mining (DMHM-PAKDD)*, pp. 36–46, 2012.

PROFESSIONAL SERVICE

- Invited Session Chair, IEEE/ACM ASONAM 2015
- Invited Reviewer of IEEE Transactions on Big Data (TBD)
- Invited Reviewer of IEEE Transactions on Knowledge and Data Engineering (TKDE)
- Invited Reviewer of IEEE Transactions on Neural Networks and Learning Systems (TNNLS)
- Invited Reviewer of SIAM Statistical Analysis and Data Mining (SAM)
- Invited Reviewer of Social Network Analysis and Mining (SNAM)
- Invited Reviewer of Science China Information Sciences (SCIS)
- External Reviewer of AAAI 2016, ICDM 2015, CIKM 2015, KDD 2015, KDD 2016, SDM 2015, SDM 2017, ASONAM 2013

SKILLS

Proficient in Python, Java, SQL, R, TensorFlow, Weka, Tableau, PHP, JavaScript, C/C++, LaTeX

Last Updated February 24, 2017