

Arindam Paul

CONTACT INFORMATION

E-mail: arindampaul.bits@gmail.com

LinkedIn: linkedin.com/in/arndmpaul/

INTERESTS

Machine Learning, Deep Learning, Natural Language Processing

PROGRAMMING SKILLS

Programming: Python, MATLAB, R

Data Science: Keras, Tensorflow, PySpark, Scikit-Learn, XGBoost, LightGBM, Gensim, NLTK, Pandas, Numpy, Seaborn, Matplotlib, Shapley, PyTorch

Web Development : HTML/Markdown, CSS, JavaScript, Ruby On Rails

EDUCATION

Northwestern University, Evanston, Illinois

Ph.D., Computer Science, Sep 2019

Advisors: Prof. Alok Choudhary, Prof. Ankit Agrawal

Northwestern University, Evanston, Illinois

Master of Science, Computer Science, Sep 2014

Birla Institute of Technology & Science, Pilani, Rajasthan India

Master of Engineering (Hons.), Software Systems, May 2012

Birla Institute of Technology & Science, Pilani, Rajasthan India

Bachelor of Engineering (Hons.), Chemical Engineering, Dec 2009 (minor in Computer Science)

PROFESSIONAL EXPERIENCE

American Family Insurance, Madison, Wisconsin

Machine Learning Scientist

Oct'19-

- Developing deep learning models for determining motor vehicle violation risk
- Designing insurance-based language models for predicting claims fraud

Northwestern Mutual Life Insurance, Milwaukee, Wisconsin

Data Science Intern

Jun '18 - Aug '18

- Developed distributed image to text conversion algorithms for detecting responses from scanned questionnaires
- Designed a noise reduction algorithm to denoise scanned and photocopied questionnaires

Boeing Cybersecurity, Sunnyvale, California

Summer Research Intern

Jun '13 - Sep '13

- Understanding Collaboration Among Online Advertising and Analytics Services
- Observed multiple 3rd-party services sharing user's information with each other
- Investigated how these services use means to obfuscate parameter sharing

RESEARCH EXPERIENCE

Northwestern University, Evanston, Illinois USA

Research Assistant

Fall '12 - present

Ensemble Nets on Mixed Representations for Chemical Property Prediction (Tensorflow, Keras) **Sep '18 -**

- Created CheMixNet - a deep neural network that combines molecular fingerprints to develop a generalized architecture for chemical property prediction
- Designed SINet - a deep network that combines two different textual representations SMILES and InChI for predicting chemical properties
- Expanded SINet for transfer learning tasks from a 2.3 million dataset to a smaller 350 compound dataset
- Developed ChemensembleNet - a deep network that combines different textual, molecular fingerprints and molecular graph representations of molecules to achieve better results than individual representations

- Deep Learning-based Predictive Model for Additive Manufacturing (Tensorflow, Keras) **Nov '16 -**
- Created Hidden Markov models for time series analysis of heat flux data
 - Investigating Recurrent Neural Network models to predict point-wise temperature information for accelerating additive manufacturing simulations
- Solar Cell Efficiency Prediction using Molecular Fingerprints (Tensorflow, Scikit Learn) **Mar '16 -**
- Designed Deep Neural Network and Random Forest models for predicting power conversion efficiency of solar cells using chemical fingerprints, and achieved mean square percentage error between 1.5-2 %
 - Designed an online application for material scientists to get an estimation of power efficiency
- Very Deep Neural Networks for Predicting Formation Stability (Tensorflow) **Mar '16 - Sept '17**
- Constructed Neural Network Models with 18-25 layers to predict formation energy of a chemical compound
 - Attained 20 % higher accuracy than the state-of-the-art models using Random Forests that would allow domain scientists to explore millions of possible compounds
- Ensemble Learning-based Guided Optimization for Aircraft Design (MATLAB, Python) **Oct '15 - Dec '17**
- Created intelligent sampling algorithms to explore the constrained search space for candidate microstructures
 - Developed Feature Ranking-based Technique for Search Space Reduction of Constrained Non-Convex Optimization
 - Achieved 100x candidate microstructures compared to state-of-the-art methods that can accelerate the design-to-experiment life-cycle
- Convolutional Neural Nets for Thematic Image Classification in Pinterest(Torch) **Oct '15 - Sep '16**
- Harnessed Association Rule Mining for thematic label curation
 - Developed ConvNet Models for hierarchical classification that led to automated image categorization based on themes
- Classification of Anonymous Posts using Recurrent Neural Networks (Tensorflow) **Jan '15 - May '16**
- Generated vectorizer models using Word2vec trained on crowd-sourced (Urban Dictionary) & psycholinguistic (LIWC) dictionaries (Gensim)
 - Attained prediction accuracy of 79.8 % and 78.1 % using LSTMs and ensemble models respectively
- Facebook Confessions & Yik Yak **Jun '14- Dec '14**
- Studied question asking about sensitive topics in anonymous forums
 - Designed a system to automatically identify and classify taboo posts in anonymous forums with good accuracy
- Learning from Ads:Reverse-engineering demographics and interests **Feb '13-May '14**
- Created synthetic user profiles with different demographic and interest features and collecting ad traffic
 - Created a model by using cross-validation which can predict user features from resulting data-set and ground-truth
 - Application of the model to cellular web-data to predict user's demographics and interests on-the-fly
- Graduate Researcher* **Spring '10 - Spring '12**
- Designing an efficient Distributed Computing Solution for Data Mining **Fall '11 -Spring '12**
- Created a Beowulf Linux(Ubuntu) cluster using OpenMPI library project
 - Implemented parallel implementation of K-means for OpenMPI
 - Bench-marked sequential and parallel OpenMPI implementations of K-means clustering algorithm
 - Compared the performance with the control Hadoop cluster

Preventing Sybil attacks in P2P systems using Psychometric Tests Fall '09 - Spring '11

- Suggested a novel approach to use Psychometric Tests (Luscher Color Test and Myers Briggs Type Indicator Test) to evaluate psychometric index of users
- Cluster nodes with similar scores and in case of a particularly high-frequency zone, we treat these nodes as suspicious and further use CAPTCHAs to remove false positives.

Software Quality Evaluation using Fuzzy Multi-Criteria Approach Spring '10- Spring '11

- Employed fuzzy ratings and weights to software attributes and proposed a comprehensive model for calculating overall software quality based on ISO/IEC 9126 model.
- Tested on multiple university softwares and one industrial application.

TEACHING
EXPERIENCE

Northwestern University, Evanston, Illinois USA

Teaching Assistant

Winter '14 -

Assisted the instructor in teaching the following undergraduate level courses. Duties included sharing of responsibilities for lectures, exams, homework assignments, grades, and office hours.

- EECS 510 Social Media Mining, Spring '17, '18 & '19
- EECS 214 Data Structures and Data Management, Fall 2015
- EECS 110 Introduction to Computer Programming (Python) , Winter & Spring 2014, Winter 2015

Guest Lecturer

Spring '16

EECS 510 Social Media Mining

- Impact of “likes” and “reactions” on social media
- Anonymity in Social Media
- Crawling and scraping the web

Instructor

Summer '16

MGLC Transferable Skills Workshop on Machine Learning

Attended by McCormick Graduate students and faculty

- What and Why of Machine Learning ?
- Algorithms
- Application
- Introduction to Deep Learning

BITS Pilani, Rajasthan, India

Teaching Assistant

Jan - May '12

Assisted the instructor in teaching the following undergraduate level courses. Duties included sharing of responsibilities for exams, homework assignments, grades and leading computer lab exercises.

- CS/IS 332 Introduction to Database Systems and Application, Spring 2012.

PUBLICATIONS:
JOURNALS

A. Paul, A. Furmanchuk, W. Liao, A. Choudhary and A. Agrawal. “**Property Prediction of Organic Donor Molecules for Photovoltaic Applications using Extremely Randomized Trees**”, *Journal of Molecular Informatics*, 2019

A. Paul, W. Liao, A. Choudhary and A. Agrawal. “**Mining Anonymous Taboo Confessions using Psycho-lingual and Crowd-Sourced Dictionaries for Emotional Wellbeing**”, *Journal of Health Informatics Research (under review)*

A. Paul, W. Liao, A. Choudhary and A. Agrawal. “**Text Translation as Data Augmentation for Neural Network Modeling of Mental Health Confessions**”, *Journal of Public Library of Science (PLOS One)* (in preparation)

A. Paul, W. Liao, A. Choudhary and A. Agrawal. “**ChemsembleNet: A generalizable, transferable architecture for predicting chemical properties using multiple representations**”, *Journal of Computational Chemistry (in preparation)*

A. Paul, P. Acar, W. Liao, A. Choudhary, V.Sundararaghavan and A. Agrawal. “**Microstructure Optimization with Constrained Design Objectives using Machine Learning-Based Feedback-Aware Data-Generation**”, *Journal of Computational Materials Science*, Apr 2019

D.Jha, L.Ward, **A. Paul**, W. Liao, A. Agrawal, A. Choudhary and C. Wolverton. “**ElemNet: Deep Learning the Chemistry of Materials From Only Elemental Composition**”, *Nature Scientific Reports*, Nov 2018

M.Mozaffar, **A. Paul**, R. Al-Bahrani, S. Wolff, A. Choudhary, A. Agrawal, K. Ehmann and J.Cao. “**Data-Driven Prediction of the High-Dimensional Thermal History in Directed Energy Deposition Processes via Recurrent Neural Networks**”, *Manufacturing Letters*, Sep 2018

A. Paul, P. Acar, R.Liu, W. Liao, A. Choudhary, V.Sundararaghavan and A. Agrawal. “**Data Sampling Schemes for Microstructure Design with Vibrational Tuning Constraints**”, *Journal of American Institute of Aeronautics and Astronautics*, Mar 2018

K Haribabu, C.Hota and **A. Paul** “**GAUR: A Method to Detect Sybil Groups in Peer-to-Peer Overlays**”, *International Journal of Grid and Utility Computing*, 2012 Vol.3 ISSN : 1741-847X
<http://dx.doi.org/10.1504/IJGUC.2012.0477655>

J.S. Challa, **A.Paul***, Y.Dada, V.Nerella, P.R. Srivastava and A.P.Singh “**Integrated Software Quality Evaluation: A Fuzzy Multi-Criteria Approach**”, *Journal of Information Processing Systems (JIPS): Korean Information Processing Society, Volume 7, Number 3 (September 2011) ISSN : 1976-913X*.
<http://dx.doi.org/10.3745/JIPS.2011.7.3.473>

* = co-first author

PUBLICATIONS: CONFERENCES

A. Paul, M. Mozaffar, Z. Yang, W. Liao, A. Choudhary, J.Cao and A. Agrawal. “**A real-time iterative approach for temperature profile prediction in additive manufacturing processes**”, *6th IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 2019

A. Paul, D.Jha, W. Liao, A. Choudhary and A. Agrawal. “**Transfer Learning Using Ensemble Neural Nets for Organic Solar Cell Screening**”, *International Joint Conference on Neural Networks*, 2019

Z.Yang, D. Jha, **A. Paul**, W. Liao, A. Choudhary and A. Agrawal. “**Generative adversarial networks with mixture density networks for inverse modeling in materials microstructural design**”, *19th IEEE International Conference on Data Mining (ICDM)* (under review)

A. Paul, D.Jha, R. Al-Bahrani, W. Liao, A. Choudhary and A. Agrawal. “**CheMixNet: Mixed DNN Architectures for Predicting Chemical Properties using Multiple Molecular Representations**”, *NIPS Workshop on Machine Learning for Molecules and Materials*, 2018

R. Liu, D. Palsetia, **A. Paul**, R. Al-Bahrani, D. Jha, W. Liao, A. Agrawal, and A. Choudhary. “**PinterNet: A Thematic Label Curation Tool for Large Image Datasets**”, *Proceedings of the Workshop on Open Science in Big Data at IEEE Bigdata Conference*, 2016.

A. Paul, A. Agrawal, W. Liao, and A. Choudhary. “**AnonyMine: Mining anonymous social media posts using psycho-lingual and crowd-sourced dictionaries**”, *Proceedings of the Workshop on Issues of Sentiment Discovery and Opinion Mining at 22nd Annual ACM Conference on Knowledge Discovery and Data Mining*, 2016.

J.Birnholtz, N.A.R. Merola, and **A. Paul**. “**Is it Weird to Still Be a Virgin?? Anonymous, Locally Targeted Questions on Facebook Confession Boards**”, *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM*, 2015.

A. Paul, Varuni G., J.S. Challa, and Y. Sharma “**HADCLEAN: A Hybrid Approach for Data Cleaning Techniques in Data Warehouses**”, *Proceedings of the IEEE International Conference on Information Retrieval and Knowledge Management(CAMP),Kuala Lumpur, March,2012*

J.S. Challa, **A.Paul***, Y. Dada, V. Nerella, and P.R. Srivastava “**Quantification of Software Quality Parameters using Fuzzy Multi-Criteria Approach**,” *Proceedings of the IEEE International Conference on Process Automation Control and Computing (PACC) 2011, Coimbatore, July, 2011*

K Haribabu, **A.Paul***, and C. Hota “**Detecting Sybils in Peer-to-Peer Overlays using Psychometric Analysis Methods**,” *Proceedings of the 25th IEEE International Conference on Advanced Information Networking and Applications(AINA)*, Singapore, March 2011

AWARDS & HONORS	<ul style="list-style-type: none"> • McCormick Dean’s Commendation Fellowship, during 6th year of PhD (2017-2018) • Predictive Science and Engineering Design Fellowship, during 5th year of PhD (2016-2017) • Segal Design Cluster Fellowship, during 3rd year of PhD (2014-2015) • Walter P. Murphy Fellowship, during 1st year of PhD (2012-2013) • BITS Pilani Merit-cum-Need Scholarship during undergraduate study • Among 10 doctoral students across Northwestern selected for summer-long Research Communication Workshop, 2016 • Best TA award for recognition of teaching excellency as Teaching Assistant for Database Systems and Applications (BITS Pilani 2012) • All India Rank 1 in BITS HDSAT (admission test for graduate programs at BITS Pilani) in Software Systems • All India Rank 64 & State Rank 9 in National Science Olympiad among more than half million participants during freshmen year of high-school
LEADERSHIP	<ul style="list-style-type: none"> • President and Treasurer , Northwestern Toastmasters Club (2015-16, 2016-17) • President and Founding Member, Northwestern Creative Writing Club • President, Northwestern University Cricket Club • Co-Facilitator, Northwestern Dialogue Group • Mentor, Brave Initiatives (http://www.braveinitiatives.com/)
SELECTED COURSE & SIDE PROJECTS	<ul style="list-style-type: none"> • Developed a Sentiment Analysis Tool to find the most interesting or controversial events at the 2014 Golden Globe Awards from user-Tweets (Python) Spring '14 • Developed a web application to track a portfolio of a user’s stocks. Used data mining techniques to analyze and predict stock and portfolio performance using historical data. (Perl, SQL) Fall '12 • Developed a real-time tool starts an alarm when a designated bus is 'x' (customizable) min away from the closest bus stop by scraping CTA bus tracker webpage (Python). Fall '14.