Prithwish Chakraborty

Data Scientist
IBM Watson Health, Cambridge, USA

January 11, 2018
prithwish.chakraborty@ibm.com
https://prithwi.github.io/

Education

Fall 2010 - Spring 2016	**PhD Student** in Computer Science, Virginia Tech, USA. **GPA:** 3.97/4.00 **Advisor** Dr. Naren Ramakrishnan
2006 - 2010	**B.E.** in Electronics and Telecomm. Engg., Jadavpur University, India. **GPA:** 9.27/10

Professional Experience

- **IBM Watson Health** Cambridge, MA
  Data Scientist. Manager: Dr. Faisal Farooq *Aug 2016 - Current*
  - Working in the RWE team as a data scientist. Designing and developing data analytic workbooks to demonstrate usage of EHR data towards health prediction. Daily responsibilities of this role include analyzing tens of millions of patient records to draw inferences and build models towards use cases such as risk prediction and disease outcome characterization using IBM Real World Evidence. These models and inferences are intended to capture many real-world effects such as the efficacy of drugs amongst demographic cohorts and total cost of treatment for diseases amongst people with varied comorbidities. More specifically, risk modeling of the diseases is aimed at identifying patient factors that capture the risk associated with severe outcomes from diseases and can lead to timely interventions and/or treatment plans. Pharmaceutical companies can use these analyses for better clinical trials and/or drug designs whereas payers can also use these to better understand the patterns of effective treatments with an overall goal towards timely and better patient care. Also involved in day-to-day development activities towards consumer driven product.
  - Involved in general research across teams towards public health forecasting
- **Virginia Tech** Arlington, VA
  GRA, Discovery Analytics Center. Advisor: Dr. Naren Ramakrishnan *Fall 2011 - Spring 2016*
  - Graduate assistant work focused on developing data driven models towards public health forecasting from multiple sources. In general, updates from public agencies are inherently delayed due to the lags in traditional disease surveillance networks. On the other hand, more real-time albeit noisy information about disease states can be found from sources such as news and social media. These models aim to capture the weakly correlated signals about diseases from these disparate sources and via data assimilation and augmentation generates forecasts about current disease state. Also, developed streaming data ingestion module for Disease surveillance, Weather and News. Implemented big-data disease forecasting pipeline for EMBERS in *python* over EMBERS *AWS* cluster framework. Forecasts were sent in real-time without human supervision and continuously evaluated
  - Worked on disease forecasting challenges and projects with several agencies such as IARPA and CDC. Collaborated with several institutes such as *NDSSL*, Virginia Tech, *Healthmap*, Harvard Medical School and *YeLab*, University of Michigan, Ann Arbor
  - Helped organize and run flu forecasting market in collaboration with Scicast, George Mason University
  - Collaborated on several projects such as pattern detection in Ionospheric Radiation, with ECE

VT and solar Photo-voltaic energy prediction with HP Labs

- **Amazon Web Services** Seattle, WA
  SDE Intern, mentored by Paul Sharpe *Summer 2011*
  - Worked on updating an Internal Console of AWS Beanstalk towards more real-time status monitoring of the same. Implemented in *java* using *Spring-Hibernate* framework using a combination of *JSP, js, css, and html5*
- **Virginia Tech** Blacksburg, VA
  GTA in Dept. of Computer Science. Instructor: Mr. N.D. Barnette *Fall 2010 - Spring 2011*
  - Was a Teaching assistant for introduction to C++ programming to a class of $\approx$ 200 students. Held office hours, 3 times a week, to tutor students on the subject, answer any questions from the lectures and help them with their homework. Also, graded assignments over the semester as well as final exam.
- **Indian Institute of Technology, Delhi** Delhi, IN
  Research Intern, advised by Prof. B.K. Panigrahi *Jun 2009 - Jun 2009*
  - Worked on application of heuristic optimization and machine learning methods to power system characterization and early Epilepsy detection using wavelets. Published several papers to several journals and conferences.

Research Statement

Research Focus: Data-driven spatio-temporal modeling under weakly correlated signals

- *Application area:* Infectious Disease modeling
- *Broad Focus:* Data Science, Machine Learning and Pattern Recognition.

Activities

Technical Talks	P. Chakraborty. Data science made easy in Jupyter notebooks using PixieDust and InsightFactory. `http://oreil.ly/2CV7CMI`, Aug 2017. Accessed: 2018-01-11
Tutorial	P. Chakraborty, M. Marathe, N. Ramakrishnan, and A. Vulikanti. Computational Epidemiology and Public Health Policy Planning, Feb 2016. Tutorial, AAAI 2016
Invited Talks	P. Chakraborty. Data Driven Model for Disease Forecasting, 10 2014. Invited Talk, First IEEE international Workshop on Big Data in Computational Biology (BCDE 2014)
Patent	M. Marwah, M. Arlitt, P. Chakraborty, and N. Ramakrishnan. Predicting near-future photovoltaic generation, Sept. 28 2012. US Patent 20140095076A1
Opinion Piece	P. Chakraborty. US Flu Forecast: Exploring links between national and regional level seasonal characteristics. `http://bit.ly/1CSHTk7`, 2014. Accessed: 2015-03-21

Publications

Current Publications

2017 P. Chakraborty, V. Gopalakrishnan, S. M. Alford, and F. Farooq. A novel data-driven framework for risk characterization and prediction from electronic medical records: A case study of renal failure. *arXiv preprint arXiv:1711.11022*, 2017

S. Ghosh, P. Chakraborty, B. L. Lewis, M. S. Majumder, E. Cohn, J. S. Brownstein, M. V. Marathe, and N. Ramakrishnan. Gell: Automatic extraction of epidemiological line lists from open sources. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1477–1485. ACM, 2017

S. Ghosh, P. Chakraborty, E. O. Nsoesie, E. Cohn, S. R. Mekaru, J. S. Brownstein, and N. Ramakrishnan. Temporal topic modeling to assess associations between news trends and infectious disease outbreaks. *Scientific reports*, 7:40841, 2017

F. S. Tabataba, P. Chakraborty, N. Ramakrishnan, S. Venkatramanan, J. Chen, B. Lewis, and M. Marathe. A framework for evaluating epidemic forecasts. *BMC infectious diseases*, 17(1):345, 2017

2016 P. Chakraborty, S. Muthiah, R. Tandon, and N. Ramakrishnan. Hierarchical Quickest Change Detection via Surrogates. *arXiv preprint arXiv:1603.09739*, 2016

S. Ghosh, P. Chakraborty, E. Cohn, J. S. Brownstein, and N. Ramakrishnan. Characterizing diseases from unstructured text: A vocabulary driven word2vec approach. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 1129–1138. ACM, 2016

H. Wu, Y. Ning, P. Chakraborty, J. Vreeken, N. Tatti, and N. Ramakrishnan. Generating Realistic Synthetic Population Datasets. *arXiv preprint arXiv:1602.06844*, 2016

2015 P. Khadivi, P. Chakraborty, R. Tandon, and N. Ramakrishnan. Time Series Forecasting via Noisy Channel Reversal. In *Machine Learning for Signal Processing (MLSP), 2015 IEEE 25th International Workshop on*, pages 1–6. IEEE, 2015

Z. Wang, P. Chakraborty, S. R. Mekaru, J. S. Brownstein, J. Ye, and N. Ramakrishnan. Dynamic Poisson Autoregression for Influenza-Like-Illness Case Count Prediction. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1285–1294. ACM, 08 2015

H. Wu, P. Chakraborty, S. Ghosh, and N. Ramakrishnan. Forecasting Influenza in Senegal with Call Detail Records, 04 2015. NETMOB 2015

2014 P. Chakraborty, P. Khadivi, B. Lewis, A. Mahendiran, J. Chen, P. Butler, E. O. Nsoesie, S. R. Mekaru, J. S. Brownstein, M. V. Marathe, and N. Ramakrishnan. Forecasting a Moving Target: Ensemble Models for ILI Case Count Predictions. In *Proceedings of the 2014 SIAM International Conference on Data Mining, Philadelphia, Pennsylvania, USA, April 24-26, 2014*, pages 262–270, 2014

2012 P. Butler, P. Chakraborty, and N. Ramakrishan. The Deshredder: A visual analytic approach to reconstructing shredded documents. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*, pages 113–122. IEEE, 2012

P. Chakraborty, M. Marwah, M. Arlitt, and N. Ramakrishnan. Fine-Grained Photovoltaic Output Prediction Using a Bayesian Ensemble. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*, page online, 2012

Other Publications

2017 P. Chakraborty, B. Lewis, S. Eubank, M. Marathe, J. Brownstein, and N. Ramakrishnan. What to know before Forecasting the Flu. to appear in PLOS Computational Biology, 2017

2012 P. Chakraborty, G. G. Roy, B. Panigrahi, R. Bansal, and A. Mohapatra. Dynamic economic dispatch using harmony search algorithm with modified differential mutation operator. *Electrical Engineering*, 94(4):197–205, 2012

T. K. Gandhi, P. Chakraborty, G. G. Roy, and B. K. Panigrahi. Discrete harmony search based expert model for epileptic seizure detection in electroencephalography. *Expert Systems with Applications*, 39(4):4055–4062, 2012

2011 G. G. Roy, S. Das, P. Chakraborty, and P. N. Suganthan. Design of non-uniform circular antenna arrays using a modified invasive weed optimization algorithm. *Antennas and Propagation, IEEE Transactions on*, 59(1):110–118, 2011

P. Chakraborty, S. Das, G. G. Roy, and A. Abraham. On convergence of the multi-objective particle swarm optimizers. *Information Sciences*, 181(8):1411–1425, 2011

2010 G. G. Roy, P. Chakraborty, and S. Das. Designing fractional-order $PI^{\lambda}D^{\mu}$ controller using differential harmony search algorithm. *International Journal of Bio-Inspired Computation*, 2(5):303–309, 2010

G. G. Roy, P. Chakraborty, S.-Z. Zhao, S. Das, and P. N. Suganthan. Artificial foraging weeds for global numerical optimization over continuous spaces. In *IEEE Congress on Evolutionary Computation*, pages 1–8, 2010

2009 P. Chakraborty, G. G. Roy, S. Das, and B. Panigrahi. On population variance and explorative power of invasive weed optimization algorithm. In *Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on*, pages 227–232. IEEE, 2009

G. G. Roy, B. Panigrahi, P. Chakraborty, and M. K. Mallick. On optimal feature selection using modified harmony search for power quality disturbance classification. In *Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on*, pages 1355–1360. IEEE, 2009

P. Chakraborty, G. G. Roy, S. Sinha, S. Bose, A. Mondal, and S. Das. Automatic Shape Independent Clustering Inspired By Ant Dynamics. In *The Proceedings of International Workshop on Machine Intelligence Research organized by Machine Intelligence Research Labs*, pages 64–74, 2009

P. Chakraborty, G. G. Roy, S. Das, D. Jain, and A. Abraham. An improved harmony search algorithm with differential mutation operator. *Fundamenta Informaticae*, 95(4):401–426, 2009

D. Jain, G. G. Roy, P. Chakraborty, and S. Das. Fuzzy Entropy-based Object Segmentation with an Inertia-Adaptive PSO. In *Advanced Computing and Communications, 2008. ADCOM 2008. 16th International Conference on*, pages 13–18. IEEE, 2008

Participation in Conferences

- **NIPS 2017**, Thirty-first Annual Conference on Neural Information Processing Systems, Long Beach, USA, December 2017.
- **AMIA 2017**, American Medical Informatics Association (AMIA) 2017 Annual Symposium, Washington DC, USA, November 2017.
- **OHDSI 2017**, 2017 Observational Health Data Sciences and Informatics (OHDSI) Symposium, Bethesda, USA, October 2017.
- **ISPOR 2017**, International Society for Pharmacoeconomics and Outcomes Research (ISPOR) 22nd Annual International Meeting, Boston, USA, May 2017.
- **AAAI 2016**, 30th AAAI Conference on Artificial Intelligence, Phoenix, USA, Feb 2016
- **KDD 2015**, 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, August 2015.
- **SDM 2014**, 2014 SIAM International Conference on Data Mining, Philadelphia, USA, April 2014.
- **DDD 2013**, 2nd International Conference on Digital Disease Detection, San Francisco, USA, September 2013.
- **AAAI 2012**, 26th AAAI Conference on Artificial Intelligence, Toronto, Canada, July 2012.

Awards and Honors

Award	Eminence and Excellence Cash Award, IBM	2017
	Manager's Choice Award for *Put the client first*, IBM	2017
	Student Travel Award, AAAI	2012
Honor	Vice President, Alpha Epsilon Lambda	2013
	Upsilon Pi Epsilon, Inducted	2013
	Phi Kappa Phi, Invited	2014
	Golden Key,Invited	2014
	Tau Beta Pi,Invited	2013
Other	President, Indian Student Association at Virginia Tech	2012-2013

Technical Skills

Programming **Python**, **C/C++**, R, Java, Matlab

Javascript, Perl, HTML/CSS

Frameworks NOSQL: MongoDB, MapReduce, Django, NodeJS, Spring