

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. DO NOT EXCEED FIVE PAGES.

NAME: Chattopadhyay, Ishanu

eRA COMMONS USER NAME (credential, e.g., agency login): ishanu

POSITION TITLE: Assistant Professor

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE (if applicable)	END DATE MM/YYYY	FIELD OF STUDY
The Pennsylvania State University, State College, PA	MS	08/2004	Mechanical Engineering
The Pennsylvania State University, State College, PA	MA	08/2005	Mathematics
The Pennsylvania State University, State College, PA	PHD	08/2006	Mechanical Engineering
The Pennsylvania State University, State College, PA	Postdoctoral Fellow	08/2010	Computer Science
Cornell University, Ithaca, NY	Postdoctoral Fellow	08/2013	Computer Science

A. Personal Statement

Professor Ishanu Chattopadhyay research focuses on large-scale data analysis, machine learning, and automated model discovery with minimal human intervention. Leading the laboratory for Zero Knowledge Discovery, Dr. Chattopadhyay is interested in unraveling complex phenomena in biology, medicine, clinical decision-making, epidemiology of complex diseases and their screening and diagnosis, by leveraging and developing algorithms for sophisticated pattern recognition and discovery. More broadly, the laboratory focuses on the design and deployment of new learning algorithms, and new extensions to existing techniques trying to understand what is theoretically possible, what can be achieved in practice, and, more generally, what are the hard limits of the data science revolution. Dr. Chattopadhyay's work resides at the cusp of several disciplines - artificial intelligence, statistical theory, formal languages, dynamical systems, and machine learning; formulating tools that work where subject matter expertise is scant; hopefully answering questions that we have not yet thought to ask.

1. Huang Y, Rotaru V, Chattopadhyay I. Sequence likelihood divergence for fast time series comparison. Knowledge and Information Systems. 2023 March 16; :- Available from: <https://link.springer.com/10.1007/s10115-023-01855-0> DOI: 10.1007/s10115-023-01855-0
2. Onishchenko D, Marlowe R, Ngufor C, Faust L, Limper A, Hunninghake G, Martinez F, Chattopadhyay I. Screening for idiopathic pulmonary fibrosis using comorbidity signatures in electronic health records. Nature Medicine. 2022 September 29; 28(10):2107-2116. Available from: <https://www.nature.com/articles/s41591-022-02010-y> DOI: 10.1038/s41591-022-02010-y
3. Rotaru V, Huang Y, Li T, Evans J, Chattopadhyay I. Event-level prediction of urban crime reveals a signature of enforcement bias in US cities. Nature Human Behaviour. 2022 June 30; 6(8):1056-1068. Available from: <https://www.nature.com/articles/s41562-022-01372-0> DOI: 10.1038/s41562-022-01372-0
4. Onishchenko D, Huang Y, van Horne J, Smith P, Msall M, Chattopadhyay I. Reduced false positives in autism screening via digital biomarkers inferred from deep comorbidity patterns. Science Advances. 2021 October 08; 7(41):- Available from:

B. Positions, Scientific Appointments and Honors

Positions and Scientific Appointments

2016 - Assistant Professor, University Of Chicago, Chicago, IL

2014 - 2016 Research Scientist, Computation Institute, University of Chicago, Chicago, IL

Honors

2020 - 2022 Young Faculty Award 2020, Defense Advanced Research Projects Agency

C. Contribution to Science

1. Analyzing complex dependencies in biosystems, and massive databases of health records is a daunting computational challenge of scale. Chattopadhyay's work in this direction proceeds with the introduction a new class of non-parametric non-linear zero-knowledge pattern inference algorithms. Classically, modeling proceeds by fixing a structure or a parameterized family, and tuning parameters against the data at hand. Data-driven nonparametric approaches with access to massive databases now allow us to explore unsupervised modeling, that make very few prior assumptions. In addition to eliminating structural biases, unsupervised inference have the potential for true pattern discovery, and finding novel and highly precise screening and diagnosis tools, that may be deployed with little additional resource demand.
 - a. Onishchenko D, Marlowe R, Ngufor C, Faust L, Limper A, Hunninghake G, Martinez F, Chattopadhyay I. Screening for idiopathic pulmonary fibrosis using comorbidity signatures in electronic health records. *Nature Medicine*. 2022 September 29; 28(10):2107-2116. Available from: <https://www.nature.com/articles/s41591-022-02010-y> DOI: 10.1038/s41591-022-02010-y
 - b. Onishchenko D, Rubin D, van Horne J, Ward R, Chattopadhyay I. Cardiac Comorbidity Risk Score: Zero-Burden Machine Learning to Improve Prediction of Postoperative Major Adverse Cardiac Events in Hip and Knee Arthroplasty. *Journal of the American Heart Association*. 2022 August 02; 11(15):- . Available from: <https://www.ahajournals.org/doi/10.1161/JAHA.121.023745> DOI: 10.1161/JAHA.121.023745
 - c. Onishchenko D, Huang Y, van Horne J, Smith PJ, Msall ME, Chattopadhyay I. Reduced false positives in autism screening via digital biomarkers inferred from deep comorbidity patterns. *Sci Adv*. 2021 Oct 8;7(41):eabf0354. PubMed Central PMCID: PMC8494294.
 - d. Jia G, Li Y, Zhang H, Chattopadhyay I, Boeck Jensen A, Blair DR, Davis L, Robinson PN, Dahlén T, Brunak S, Benson M, Edgren G, Cox NJ, Gao X, Rzhetsky A. Estimating heritability and genetic correlations from large health datasets in the absence of genetic data. *Nat Commun*. 2019 Dec 3;10(1):5508. PubMed Central PMCID: PMC6890770.
2. Dr. Chattopadhyay's work on the inverse Gillespie algorithm iGillespie and GenESeSS provide two complementary approaches to de novo modeling in biosystems. In the former we can back out complex regulatory circuits, at various abstraction levels from single cell expression data, and in the latter we can infer generative stochastic models from discrete time series. Dr. Chattopadhyay's other key contribution is the Data Smashing algorithm. Most "data mining" algorithms today rely on a human expert to specify what "features" of the data are relevant. Relying on experts is error prone, and unlikely to keep pace with the growing complexity of biomedical data. Data Smashing provides a universal way to circumvent the reliance on human experts, quantifying the similarity between data streams without prior knowledge of where they were generated, how they are encoded, and what they represent; in essence creating a universal quantification of similarity, in the same sense that Kolmogorov complexity is a universal quantification of complexity. Data Smashing allows detection of subtle changes in the underlying stochastic dynamics driving biochemical processes, and provide

crucial building blocks for personalized precision medicine: applications range from isolating cohorts of patients that respond similarly to specific therapies, to detecting abnormal cardiac rhythms, to identifying distinct classes of epileptic pathologies from EEG recordings, to classifying individual microbiomes.

- a. Rotaru V, Huang Y, Li T, Evans J, Chattopadhyay I. Event-level prediction of urban crime reveals a signature of enforcement bias in US cities. *Nature Human Behaviour*. 2022 June 30; 6(8):1056-1068. Available from: <https://www.nature.com/articles/s41562-022-01372-0> DOI: 10.1038/s41562-022-01372-0
 - b. Chattopadhyay I, Lipson H. Data smashing: uncovering lurking order in data. *Journal of The Royal Society Interface*. 2014 December 06; 11(101):20140826-. Available from: <https://royalsocietypublishing.org/doi/10.1098/rsif.2014.0826> DOI: 10.1098/rsif.2014.0826
 - c. Chattopadhyay I. Scalable ϵ -Optimal Decision-Making and Stochastic Routing in Large Networks via Distributed Supervision of Probabilistic Automata. *SIAM Journal on Control and Optimization*. 2014 January; 52(4):2512-2547. Available from: <http://epubs.siam.org/doi/10.1137/110857507> DOI: 10.1137/110857507
 - d. Chattopadhyay I, Kuchina A, Suel G, Lipson H. Inverse Gillespie for inferring stochastic reaction mechanisms from intermittent samples. *Proceedings of the National Academy of Sciences*. 2013 July 22; 110(32):12990-12995. Available from: <http://www.pnas.org/cgi/doi/10.1073/pnas.1214559110> DOI: 10.1073/pnas.1214559110
3. Algorithmic Analysis of Genomic Viral Genome Databases at scale to predict sequence divergence and emergence of novel pathogens: As we begin to recover from the COVID-19 pandemic, a key question is if we can avert such disasters in future. Current surveillance protocols generally focus on qualitative impact assessments of viral diversity. These efforts are primarily aimed at ecosystem and human impact monitoring, and do not help to precisely quantify emergence. Dr. Chattopadhyay's recent and ongoing work aims to address this gap in knowledge by designing new scalable algorithms that carry out pattern discovery in sequence databases, with the objective of predicting spillover and species jumps before they happen. A new metric for comparing biologically meaningful genomic differences, the q-distance, precisely quantifies the probability of spontaneous jump by random chance. Learning from patterns of mutations from large sequence databases, the q-distance adapts to the specific organism, the background population, and realistic selection pressures; demonstrably improving inference of ancestral relationships and future trajectories. As important application, Dr. Chattopadhyay's work shows that the q-distance predicts future strains for seasonal Influenza, outperforming World Health Organization (WHO) recommended flu-shot composition almost consistently over two decades.
- a. Chattopadhyay, Ishanu,, Sizemore, Nicholas,, Oliphant, Kaitlyn,, Martin, Camilia,, Claud, Erika,. A Digital Twin of the Infant Microbiome to Predict Neurodevelopmental Deficits. [Preprint]. 2023 January 05. DOI: 10.21203/rs.3.rs-2406518/v1
 - b. Chattopadhyay, Ishanu,, Wu, Kevin,, Li, Jin,, Esser-Kahn, Aaron,. Emergenet: Fast Scalable Pandemic Risk Assessment of Influenza A Strains Circulating In Non-human Hosts. [Preprint]. 2022 December 23. DOI: 10.21203/rs.3.rs-2336091/v1
 - c. Huang Y, Chattopadhyay I. Universal risk phenotype of US counties for flu-like transmission to improve county-specific COVID-19 incidence forecasts. *PLOS Computational Biology*. 2021 October 14; 17(10):e1009363-. Available from: <https://dx.plos.org/10.1371/journal.pcbi.1009363> DOI: 10.1371/journal.pcbi.1009363
 - d. Chattopadhyay I, Kiciman E, Elliott JW, Shaman JL, Rzhetsky A. Conjunction of factors triggering waves of seasonal influenza. *Elife*. 2018 Feb 27; 7 PubMed Central PMCID: PMC5864297.

Complete List of Published Work in My Bibliography:

<https://www.ncbi.nlm.nih.gov/myncbi/ishanu.chattopadhyay.2/bibliography/public/>