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 (Medicine)

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)* | Completion Date  MM/YYYY | FIELD OF STUDY |
| Jadavpur University, Kolkata, India | B.S. | 06/2001 | Mechanical Engineering |
| The Pennsylvania State University, State College, Pennsylvania USA | M.S. | 12/2005 | Mechanical Engineering |
| The Pennsylvania State University, State College, Pennsylvania USA | M.A. | 05/2006 | Mathematics |
| The Pennsylvania State University, State College, Pennsylvania USA | Ph.D. | 08/2006 | Mechanical Engineering |
| The Pennsylvania State University, State College, Pennsylvania USA | Postdoctoral Scholar | 08/2011 | Mechanical Engineering |
| Cornell University, Ithaca, New York USA | Postdoctoral Scholar | 08/2013 | Computer Science |
|  |  |  |  |

**A. Personal Statement**

Professor Ishanu Chattopadhyay research focuses on the core algorithmic principles in large-scale data analysis with minimal human intervention, or where there is little prior domain expertise. Leading the laboratory for Zero Knowledge Discovery, Dr. Chattopadhyay is interested in unravelling complex phenomena in biology, biomedicine, clinical decision-making, epidemiology of complex diseases, and human social interactions. More broadly, the laboratory focuses on the design and exploration 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 revolution in data science that we are currently experiencing. 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 with little prior experience, and hopefully answering questions that we have not yet thought to ask.

Dr. Chattopadhyay’s background in machine learning is particularly suited for the modeling, inference, prediction and simulation of microbial dynamics in the gut environment, and distilling insights on how such dynamical patterns relate to, and possibly drive, phenotypic outcomes and clinical prognoses for infants with premature births. Leveraging his prior work on data driven discovery of generative models, and inference of epidemiological dynamics of viral infections, Dr. Chattopadhyay’s laboratory has begun working with Dr. Erika Claud’s group analyzing preliminary microbiome datasets from the current effort, and is beginning to distill significant patterns of dynamical interactions with potentially important clinical significance. However, the full scope of the machine inference algorithms developed by the Chattopadhyay lab will be applicable once more data from more subjects revealing richer sets of dynamical variations becomes available in the course of the proposed effort.

**B. Positions and Honors**

**Professional Experience**

07/2014-08/2016 Research Scientist, Computation Institute, Institute for Genomics & Systems Biology,

University of Chicago, Chicago, IL

09/2016- Assistant Professor, Department of Medicine, Section of Hospital Medicine, University of Chicago, Chicago, IL

**Honors and Awards**

2010 Best Paper In Session “Learning-Control” American Control Conference 2010

I. Chattopadhyay, Y. Wen and A. Ray, Pattern Classification In Symbolic Streams Via Semantic Annihilation of Information, American Control Conference, 2010, Baltimore, MD, June 30-July 2

2009 Best Paper In Session “Path-Planning” American Control Conference 2009

I. Chattopadhyay and A. Ray, Optimal Path-Planning under Finite Memory Obstacle Dynamics Based on Probabilistic Finite State Automata Models

2009 Best Paper In Session “Large Scale Systems” American Control Conference 2009

I. Chattopadhyay and A. Ray, Supervised Self-Organization of Large Homogeneous Swarms Using Ergodic Projections of Markov Chains

2009 Best Paper In Session “Agent Based Systems II” American Control Conference 2009

W. Lu, I. Chattopadhyay, G. Mallapragada and A. Ray, A Real Time Implementable All-Pair Dyn. Planning Algo. for Robot Nav. Based on the Renorm. Measure of Prob. Reg. Languages

**C. Contributions to Science**

1. Analyzing complex dependencies in biosystems, and massive databases of health records is a daunting computational challenge of scale. Dr. 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, particularly ones that we have not thought of to test for. Dr. Chattopadhyay’s work on the inverse Gillespie algorithm iGillespie and GenESeSS provide two complementary approaches to such modeling. 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.
2. Chattopadhyay, E. Kiciman, J. Elliott, J. Shaman and A. Rzhetsky , *"Conjunction of factors triggering waves of seasonal influenza"*, Elife, eLife 2018;7:e30756, 2018
3. Chattopadhyay, A. Kuchina, G. M. Suel, and H. Lipson, “Inverse gillespie for inferring stochastic reaction mechanisms from intermittent samples,” Proceedings of the National Academy of Sciences, vol. 110, no. 32, pp. 12 990–12 995, 2013.
4. Chattopadhyay and H. Lipson, “Abductive learning of quantized stochastic processes with probabilistic finite automata,” Philos Trans A Math Phys Eng Sci, vol. 371, no. 1984, p. 20110543, 2013.
5. ——, “Data smashing: uncovering lurking order in data,” Journal of The Royal Society Interface, vol. 11, no. 101, 2014. [Online]. Available: <http://rsif.royalsocietypublishing.org/content/11/101/20140826>
6. Jia, G., Li, Y., Zhang, H., Chattopadhyay I, et al. Estimating heritability and genetic correlations from large health datasets in the absence of genetic data. Nat Commun 10, 5508 (2019). https://doi.org/10.1038/s41467-019-13455-0

**D. Research Support**

**Ongoing Research Support**

Defense Advanced Research Projects Agency/ATI (Weinberger) 10/01/18-03/31/22

*PREEMPT: Platform Technologies for Enzootic Disease Control*

The main goal of this proposal is to develop models of MHV-encoded transmissible recombinant vaccine (TRV) control of CCHFV.

Role: Co-PI

Defense Advanced Research Projects Agency (Chattopadhyay) 09/24/18-03/24/20

*Q-Nets: Accelerated Robust Learning Via Deep Knowledge Integration*

The main goal of this proposal is to develop a rapid and autonomous ML-based therapeutic prediction platform capable of learning the druggable features of any viral pathogen, from its genome alone and from limited laboratory perturbations (i.e., mutagenesis) of that genome.

Role: PI

Defense Advanced Research Projects Agency (Chattopadhyay) 01/01/17-12/31/20

*ZeD: Zero-Knowledge Discovery Using Data Smashing*

The main goal of this proposal is designed to address zero knowledge inference: the task of finding models from raw data when we do not necessarily know what the correct model structure is. In addition to designing primitives to discover generative models from raw data, we will address the featurization issuerigorously.

Role: PI

**Completed Research Support**

Neubauer Collegium: Faculty Initiated Research 07/01/17-06/30/19

Interrogating Free Will: Zero-knowledge Machine Inference To Predict, Intervene, Mitigate & Make sense of Criminal Infractions & Its Ethical Implications

Role: PI