ARUN NAMPALLY

4311 Norwalk Dr◇ Apt T-302 ◇ San Jose, CA 95129
+1-(631)-538-5071 ◇ arun.nampally@invitae.com

EDUCATION

Stony Brook University, Stony Brook Jan 2011 - Aug 2018
PhD in Computer Science, GPA: 3.91/4
Thesis: Symbolic Inference Techniques in Probabilistic Logic Programming.
Advisor: C. R. Ramakrishnan

University of Hyderabad, Hyderabad, India Aug 2005 - Aug 2007
M. Tech in Information Technology, First class with distinction.

Jawaharlal Nehru Technological University, Hyderabad, India Aug 2000 - Apr 2004
B. Tech in Computer Science, First class.

TECHNICAL SKILLS

Programming Languages	Python, C, Prolog, Java, C++
Probabilistic Programming Languages	Pyro, ProbLog
Machine Learning	PyTorch, scikit-learn, Tensorflow
Databases	MySQL, Oracle
Testing	pytest
CI-CD	Git, SVN, CVS, Jenkins, Docker
Scripting Languages	Shell scripting, Expect scripting

EXPERIENCE

Invitae Mar 2019 -
AI Scientist *San Francisco, CA*

Variant Interpretation: Worked on research and development of probabilistic models for variant interpretation using probabilistic programming and graphical models. Provided technial leadership for the development of the system called Holmes that implements a probabilistic classification system for variant interpretation.

Probabilistic Programming: Worked as lead developer for a library called *pearl* that provides a convenient wrapper around the Pyro programming language to enable users to define and train implicit graphical models.

Low-Coverage Imputation: Worked on research and development of a novel algorithm to infer fine grained Haplotype reference panel from internal sequence data using Bayesian non-parametric modeling. This allows the use of inexpensive low coverage genome sequencing to be used for variant calling and other genomic analysis tasks.

Software Engineering: Worked on a various software development and engineering tasks related to the R&D efforts. These include instrumentation of a Django web-app to collect object changes in Kafka streams, development of Docker images, CICD using Jenkins, writing test-suites, data collection and pre-processing etc.

Stony Brook University Jan 2013 - Dec 2018
Research Assistant *Stony Brook, NY*

Probabilistic Logic Programming: Worked on developing scalable inference procedures in systems combining logical and probabilistic models. This work is part of the broader effort in AI to come up with machine learning frameworks that are more declarative, expressive, interpretable and explainable. Specifically, worked in the area of *Probabilistic Logic Programming* and developed symbolic and

approximate inference techniques using ideas from constraint processing, exact and approximate inference in probabilistic machine learning (e.g. likelihood weighting, Markov Chain Monte Carlo) and reinforcement learning. Links to prototype implementations:

- **Px**: Declarative probabilistic programming system.
- Ordered symbolic derivation diagrams.
- Adaptive Markov Chain Monte Carlo inference.

Stony Brook University Jan 2011 - Jan 2013
Teaching Assistant *Stony Brook, NY*

Managed large classes (100+) as teaching assistant by conducting recitations, grading, lab sessions and office hours. *Courses*: (*i*) C programming, (*ii*) Algorithms and Data Structures, and (*iii*) Network Programming.

Teradata Jun 2010 - Jan 2011
Software Engineer *Hyderabad, India*

Teradata RDBMS: Involved in design and development of a stress test framework ("Quantum Instrumentation"), which provides test coverage for Teradata RDBMS product. Developed framework to auto-generate simulated asynchronous failure events (asynchronous aborts and query rewrites) for increasing the overall test coverage. The software was built using Java, Perl and SQL.

IBM Aug 2007 - Jan 2010
Software Engineer *Bangalore, India*

Dynamic tasking prototype of IBM POE: Involved in design and development of dynamic tasking prototype for IBM parallel operating environment (POE). Lead the effort in design of dynamic tasking feature for Message Passing Interface (MPI) 2.0 and developed resource management architecture (an iterative TCP/IP server), which arbitrates requests for resource usage by MPI processes. This software was developed using C and MPI library.

SNAP genetic programming toolkit: Involved in development of functional module to perform scan-chain optimization using SNAP programming tool kit. Conceptually, a "multi-depot vehicle routing" problem (NP-complete), which is suitable to be solved by genetic algorithms. System developed using C++.

PUBLICATIONS

Arun Nampally, Eugene Palovcak, Garrett Bernstein, Matthew Davis: **Graphical Models For Rare Sequence Variant Interpretation**. *AAAI 2021 Workshop: Trustworthy AI for Healthcare*

Arun Nampally, Timothy Zhang, C. R. Ramakrishnan: **Constraint Based Inference in Probabilistic Logic Programs (extension of PLP 2015 work)**. *Theory and Practice of Logic Programming, 18(3-4), 638-655.*

Arun Nampally, C. R. Ramakrishnan: **Inference in Probabilistic Logic programs using Lifted Explanations**. *Technical Communications of the 32nd International Conference on Logic Programming (ICLP 2016).*

Arun Nampally, C. R. Ramakrishnan: **Constraint Based Inference in Probabilistic Logic Programs**. *Proceedings of the 2nd International Workshop on Probabilistic Logic Programming co-located with 31st International Conference on Logic Programming (ICLP 2015).*

Arun Nampally, C. R. Ramakrishnan: **Adaptive MCMC-Based Inference in Probabilistic Logic Programs**. *Technical Communications of the 30th International Conference on Logic Programming (ICLP 2014).*

Nigib Sharma, Nampally Arun, Vadlamani Ravi: **An ant colony optimisation and Nelder-Mead simplex hybrid algorithm for training neural networks: an application to bankruptcy prediction in banks**. IJIDS 5(2): 188-203 (2013)