Suyash Shringarpure

Graduate Student
Carnegie Mellon University
Pittsburgh, PA

Doherty Hall 4301A
5000 Forbes Avenue
Pittsburgh PA 15213-3891
suyash@cs.cmu.edu

RESEARCH INTERESTS

Machine Learning, Computational Biology

EDUCATION

Carnegie Mellon University, Pittsburgh, PA
Third-year PhD student, Machine Learning Department.
GPA: 3.86/4.0

Indian Institute of Technology Bombay, Mumbai, India
Bachelor of Computer Science and Engineering, 2006.
GPA: 9.14/10.

WEBSITE

http://www.cs.cmu.edu/~suyash

SKILLS

Computer languages: Java, C++.
Scripting languages: Perl, Python.
Operating systems: Linux, Windows.
Mathematical software: Matlab, R.

RELEVANT COURSES

2006-08: Statistical Machine Learning, Computational Genomics, Machine Learning, Intermediate Statistics, Multimedia Databases, Probabilistic Graphical Models, Evolution.
2002-06: Hypertext Retrieval & Mining, Graphical models, Data Mining, Design & analysis of algorithms, Data Structures, Linear algebra.

EXPERIENCE

PhD Research
Advisor: Eric Xing

My research involves analyzing genetic data with the help of computational tools. I am interested in the application of machine learning methods to study genetic phenomenon such as mutation, recombination, drift and selection, which affect human evolutionary history. Some related projects are:

1. mStruct - Admixture model for genetic data incorporating mutation effects.
We developed a model to detect population structure in multi-locus genotype data in the presence of admixture and allele mutations. It has the advantage of modeling ancestral populations and also providing information about allele mutations. By capturing similarity between alleles, we are better able to model the similarities and differences between individuals from various populations.

2. CSMET: Comparative Genomic Motif Detection via Multi-Resolution Phylogenetic Shadowing.

 CSMET is a generative model for modeling functional turnover in transcription factor binding sites using phylogenetic information at the sequence and functional levels. We compared the performance of CSMET to other existing methods on Drosophila data, and observed favorable results. We were also able to detect motifs which were previously undetected by computational methods but were known in literature.

Mixed-membership models for text data:
Advisor: Steve Fienberg

 On text data, my work involved experiments for model selection in a mixed-membership model on a large dataset of articles from the PNAS collection. We also developed an extension to existing models that enforced chronology constraints in the reference model.

Senior Undergraduate Thesis: Graphical Models for Active Deduplication & Collective Type Labeling.
Guide: Soumen Chakrabarti, IIT Bombay

 We explored the application of graphical models for citation deduplication, with the aim of providing a mathematical foundation to the methods. We also observed performance improvements obtained by the use active learning methods.

PUBLICATIONS S. Shringarpure and E. P. Xing, **mStruct: A New Admixture Model for Inference of Population Structure in Light of Both Genetic Admixing and Allele Mutations**, *Genetics (in press)*.

P. Ray, S. Shringarpure, M. Kolar and E. P. Xing, **CSMET: Comparative Genomic Motif Detection via Multi-Resolution Phylogenetic Shadowing**, *PLoS Computational Biology (2008)*.

ACTIVITIES
- Vice President of the Indian Graduate Student Association at Carnegie Mellon University, 2008-Present.
- General Secretary of the Indian Graduate Student Association at Carnegie Mellon University, 2007-08.
- Department General Secretary of the Computer Science and Engineering Department, IIT Bombay, 2005-06.

AWARDS
- Designated a Presidential Fellow in the Life Sciences (2008).

REFERENCES
- Will be provided on request