Suyash Shringarpure

CONTACT INFORMATION

8013 Gates-Hillman Center — 412 651 3099
Machine Learning Department, Carnegie Mellon University — suyash@cs.cmu.edu
5000 Forbes Avenue, Pittsburgh, PA, USA - 15213 — http://www.cs.cmu.edu/~suyash/

RESEARCH INTERESTS

I am interested in developing statistical methods using machine learning for problems in computational biology and population genetics in particular.

Keywords: Machine Learning, Computational Biology, Population Genetics.

EDUCATION

Carnegie Mellon University, Pittsburgh, PA, USA

Ph.D. in Machine Learning — **2006 – present**

- Dissertation: Statistical methods for studying genetic variation in populations.
- Advisor: Eric P. Xing
- Expected graduation date: May 2012

Indian Institute of Technology Bombay, Mumbai, India

B.Tech. Computer Science and Engineering — **2002 – 2006**

AWARDS

Richard Mellon Presidential Fellowship in Life Sciences — **2008**

CMU Machine Learning Graduate Fellowship — **2006**

PUBLICATIONS

S. Shringarpure, D. Won and E. P. Xing (2011), "StructHDP: Automatic inference of number of clusters from admixed genotype data". *Bioinformatics*, Vol. 27, No. 13, Pages 324-332.

E. Airoldi, E. Erosheva, S. Fienberg, C. Joutard, T. Love, & **S. Shringarpure** (2010), "Reconceptualizing the classification of PNAS articles". *Proceedings of the National Academy of Sciences*, Vol. 107, No. 49, Pages 20899-20904.

S. Shringarpure, E. P. Xing (2009), "mStruct: Inference of population structure in light of both genetic admixing and allele mutations". *Genetics*, Vol. 182, Pages 575-593.

S. Shringarpure and E. P. Xing (2008) "mStruct: A New Admixture Model for Inference of Population Structure in Light of Both Genetic Admixing and Allele Mutations". *Proceedings of the 25th International Conference on Machine Learning* (ICML 2008).

P. Ray*, **S. Shringarpure***, M. Kolar and E. P. Xing (2008), "CSMET: Comparative Genomic Motif Detection via Multi-Resolution Phylogenetic Shadowing". *PLoS Computational Biology*, Vol. 4 No. 6, June 2008. (* indicates joint first authors)

S. Shringarpure and E. P. Xing (2012), "Mixed-membership models in population genetics". In Airoldi, Blei, Erosheva, Fienberg, Bokalders (Eds.), Handbook on Mixed Membership Models. Chapman & Hall. (In preparation, draft available on request)

EXPERIENCE

Carnegie Mellon University

Graduate Research Assistant — **2006-Present**

Research on applying machine learning methods for (a) population stratification taking evolutionary processes into account (b) identifying genes responsible for diseases using artificial selection methods in fruit flies.

Cold Spring Harbor Laboratory

Summer Internship — **June 2008-September 2008**

Worked with Mickey Atwal and Bud Mishra (NYU) on statistical methods for detecting co-selection in human populations from HapMap.

Teaching Experience

CMU Machine Learning10-701/15-781

Teaching Assistant **Fall 2008, Fall 2011**

- Core-curriculum course for SCS graduate students focusing on fundamental algorithms and theory for statistical machine learning.
- I was the Head TA for the course offering in 2011. I advised and graded 15 student projects.
- Other duties included conducting recitations, office hours, designing and grading homeworks and exams.

CMU Computational Genomics 02-710/10-810

Guest Lecturer **Spring 2009**

CMU Technights: Researching your Family Tree **May 2010**

- Creative Technology Nights for Girls is a program by Women@SCS focused on exposing middle school girls to creative technologies.
- I made a presentation to teach middle school girls about the concepts of genetic ancestry and how ancestry can be visualized with a demo of ancestry representations.
- I organized a hands-on-activity involving the representation of ancestry for the students.

Advising

Daegun Won (BS CS '10) **2009**
(Co-advised with Eric Xing)
Inference of Population Structure with Optimal Number of Ancestral Groups

Service

International Society for Molecular Biology

Student member **2009 – Present**

Reviewer

External reviewer for conferences and journals **2009 – Present**

Conferences: ISMB, RECOMB, APBC
Journals: American Association of Cancer Research, Genetics

Invited Talks

Automatic inference of number of clusters from admixed genotype data **July 2011**
International Society for Molecular Biology, Vienna, Austria

mStruct: Structure under Mutations **April 2009**
DIMACS Workshop on Algorithmics in Human Population-Genomics, Piscataway, NJ

mStruct: A New Admixture Model for Inference of Population Structure. **July 2008**
International Conference on Machine Learning, Helsinki, Finland

Research Summary

Population Stratification

Stratification of populations is important in genome-wide association studies and for understanding the evolutionary history of populations. My interests are in developing novel statistical methods to obtain accurate stratification of a given sample of individuals.

For this purpose, I have developed *mStruct*, which is a method that models admixture of populations while taking into account the allele mutations that may be present in the populations. The method uses mixed-membership models and a novel algorithm to learn the model parameters from data.

Another problem often faced when attempting stratification is to identify how many ancestral populations are required to accurately capture the stratification present in a sample. To address this problem, I have developed *StructHDP*. It infers the number of ancestral populations that best represent the stratification present in a sample and identifies the stratification simultaneously.

Artificial selection methods in association

Genome-wide association studies using the case-control design have proved to be of limited utility in shedding light on the genetic basis of multifactorial traits. I examined the utility of using artificial selection with *Drosophila Melanogaster* to improve our understanding of the genetic factors affecting complex traits. My experiments with simulated and semi-simulated data show that data generated from artificial selection allows better detection of genetic effects than ordinary case-control data using conventional association methods. Ongoing work examines whether such data allows better detection of complex effects like epistasis.

Detecting Recent Co-selection

Current work has found evidence for recent selection (upto 25,000-50,000 years ago) in many modern populations. Since genes are expressed as part of genetic interaction networks, I examined the hypothesis that selection effects act on multiple genes simultaneously. I examined the HapMap Phase 3 populations for evidence of recent co-selection over SNP pairs using an entropy-based method.

REFEREES

Eric P. Xing
Professor
Carnegie Mellon University
Pittsburgh, PA, USA
phone: 412-268-2559
e-mail: epxing@cs.cmu.edu

Stephen E. Fienberg
Professor
Carnegie Mellon University
Pittsburgh, PA, USA
phone: 412-268-2723
e-mail: fienberg@stat.cmu.edu

Kathryn Roeder
Professor
Carnegie Mellon University
Pittsburgh, PA, USA
phone: 412-268-2513
e-mail: roeder@stat.cmu.edu