Dhawal Joharapurkar

Contact Information

485, Engineering 2
Computer Science Department
University of California, Santa Cruz
Santa Cruz 95060

Phone: (831) 346-8702
email: dhawal@ucsc.edu
web: https://dhawaljoh.github.io

Education

University of California, Santa Cruz — 2016 - Present
Ph.D. in Computer Science
Advisor: Prof. Lise Getoor

Manipal Institute of Technology, Manipal — 2011 - 2015
B.Tech in Computer Science & Engineering
Bachelor Thesis: Temporal Scoping of Facts in a Knowledge Base

Honors

Regents' Fellowship, UC Santa Cruz

Work Experience

Xerox Research Centre India, Bangalore — June '15–Jan '16
Research Intern, Machine Learning and Statistics Group
Advised by Dr. Vaibhav Rajan and Sumit Negi

- Analysis of clinical notes of ICU patients using Topic Models
- Created a plotting tool to highlight different statistics across classes

Working on predictive healthcare using unstructured nursing notes.

Indian Institute of Science, Bangalore — December '14–May '15
Project Trainee, Supercomputer Education and Research Centre (SERC)
Advised by Prof. Partha Talukdar

- Temporal scoping and ordering of relations in a knowledge base
- Entity linking and disambiguation in large text corpora

Indian Institute of Technology, Kharagpur — May '14–June '14
Summer Research Intern, Dept. of Computer Science & Engineering
Advised by Prof. Sudeshna Sarkar

- Automatic profiling of Driver Behaviour on a GPS dataset provided by MHRD.
- Implemented DBSCAN algorithm to find traffic stoppage points and segmented roads based on their speed profiles
- Modified "simplekml" Python module to plot the GPS points on Google Maps

DataWeave Software Pvt. Ltd., Bangalore — May '13–Jun '13
Summer Intern

- Created data crawlers using Python that aggregated and stored content in JSON dumps
- Content available via APIs, a few listed at http://dataweave.in/apis

Online Courses

The Data Scientist's Toolkit — June '14
coursera.org, 100% — Johns Hopkins University

Design and Analysis of Algorithms — May '14
Massively Empowered Classrooms, 100% — Microsoft Research

Algorithms: Design and Analysis, Part 1 July '13
coursera.org, 98% Stanford University

Machine Learning April '14
coursera.org, 100% Stanford University

Projects

Detecting Fibrous Regions in Protein Sequences November '13 – May '14
Guide: Dr. Smitha Nair Manipal Institute of Technology, Manipal
Worked on the detection of fibrous regions in protein sequences using Support Vector Machines and Bee Colony Optimization for PCA.

Photo Tagger: Multi-class classification March '14
Rank: 84 out of 644 CSA, IISc, Bangalore
Used SVM to classify photos into various classes (people, cars, shoes, buildings, flowers). The parameters of the SVM were optimized using GridSearchCV. The features were extracted using the SIFT algorithm.

Craigslist Post Classification October '13
Accuracy: 81% Manipal Institute of Technology, Manipal
Used bag of words model, tf-idf and SVM to classify posts on Craigslist into sections based on the product description. The open dataset was available on HackerRank

Publications

1. **Online Adspace Posts' Category Classification**
   Dhawal Joharapurkar, Vaishak Salin, Vishal Krishna
   12th International Conference on Natural Language Processing, December, 2015

Workshops

International Institute of Information Technology, Hyderabad July '15
One of 60 invited participants at IASNLP - 2015 across India
Project mentor:Prof. Manish Shrivastava

- Question category classification on TREC data using topic composition features generated by LDA
- Improved area under ROC curve of existing system developed by students at IIIT-H by 3%

Talks

From Big Text to Big Knowledge Feb '15
SERC Open Day 2015 IISc, Bangalore

Skills

- **Languages:** C/C++, Python, Octave, SQL, LaTeX
- **Operating Systems:** Linux(various distributions), Microsoft Windows
- **Tools:** Emacs, Sublime Text, Enthought Canopy, IPython
- **Version Control System:** git