

**Umaa Rebbapragada**  
urebbapr@cs.tufts.edu  
<http://www.cs.tufts.edu/~urebbapr>

## EDUCATION

---

### **Tufts University**

Ph.D. Candidate in Computer Science	2006 – (May 2010)	GPA: 3.8
M.S. Computer Science	2004 – 2006	GPA: 3.8
Post-baccalaureate Minor in Computer Science (enrolled part-time while working full-time at CNET Networks)	2002 – 2004	GPA: 3.9

### **University of California, Berkeley**

B.A. Mathematics	1993 – 1997	GPA: 3.4
------------------	-------------	----------

## RESEARCH

---

**Tufts University**, *Research Assistant* to Professor Carla Brodley [9/04 – present]  
Specialization: machine learning, data mining, time series data, anomaly detection

*Class Noise Mitigation*: My thesis focuses on the detection and mitigation of label noise in training data sets. It includes a comprehensive assessment of past techniques for discarding and correcting instances suspected of having noisy labels, and a novel method for mitigating noise by using the output from the EM algorithm to calculate a probability distribution over the class labels for each example. The second aspect of my research examines how one can use that probability distribution to weight the instances during training and achieve superior results over methods that do 0,1 discarding or correcting. Finally, my research explores the implication of using label noise mitigation and instance weights to improve the performance of other supervised learning methods (e.g., co-training).

*Anomaly Detection in Time Series Data*: My method discovers both local and global anomalies from large sets of unsynchronized periodic time series data. My primary application area is astrophysics with the goal of discovering anomalies that may be indicative of a new subclass of stars. This research is done in collaboration with scientists at the Harvard-Smithsonian Center for Astrophysics.

**Jet Propulsion Laboratory**, *Intern* under Dr. Kiri Wagstaff, [6/08 – 8/08]

*Collaborative Learning*: Collaborative learning is a new learning framework developed for distributed learning environments (e.g., sensor networks) and combines the strengths of active learning, co-training and ensemble learning. We developed collaborative learning for our motivating application: classifying volcanic activity from seismic data collected from a sensor network at the Mount Erebus Volcanic Observatory.

## PUBLICATIONS

---

**U. Rebbapragada**, P. Protopapas, C. E. Brodley, and C. Alcock (2009) “Anomaly Detection in Catalogs of Periodic Variable Stars”, In *Proceedings of the 18<sup>th</sup> Astronomical Data Analysis Software and Systems Conference*.

L. Mandrake, K. Wagstaff, D. Gleeson, **U. Rebbapragada**, D. Tran, R. Castaño, S. Chien, and R. Pappalardo (2009) “Onboard Detection of Naturally Occurring Sulfur Compounds on the Surface of a Glacier using an SVM and the Hyperion Multi-Spectral Instrument”. In *Proceedings of the 2009 IEEE Aerospace Conference*.

**U. Rebbapragada**, L. Mandrake, K. Wagstaff, D. Gleeson, R. Castaño, S. Chien, and C. E. Brodley (2009) “Improving Onboard Analysis of Hyperion Images by Filtering Mislabeled Training Data Examples”. In *Proceedings of the 2009 IEEE Aerospace Conference*.

**U. Rebbapragada**, P. Protopapas, C. E. Brodley, and C. Alcock (2009) “Finding Anomalous Periodic Time Series”, *Machine Learning Journal*, Vol. 74, Issue 3, p. 281

**U. Rebbapragada**, R. Lomasky, C. E. Brodley and M. Friedl (2008) “Generating High-Quality Training Data for Automated Land-Cover Mapping”. In *Proceedings of the 2008 IEEE International Geoscience and Remote Science Symposium*.

D. Yankov, E. Keogh, and **U. Rebbapragada** (2007) “Disk Aware Discord Discovery: Finding Unusual Time Series in Terabyte Sized Datasets”. In *Proceedings of the 7<sup>th</sup> IEEE International Conference on Data Mining* [Best Paper Award]

**U. Rebbapragada** and C. E. Brodley (2007) “Class Noise Mitigation through Instance Weighting”. In *Proceedings of the 18<sup>th</sup> European Conference on Machine Learning*.

---

#### TALKS/PRESENTATIONS

---

Tips, Tricks and Software for Keeping Research Organized [10/2/2009]  
Birds of a Feather Session (with Audrey Girouard and Kristina Winbladh)  
Grace Hopper Conference 2009, Tucson, AZ

Class Noise Detection through Instance Weighting [10/1/2008]  
Ph.D. Forum on Artificial Intelligence and Learning Systems  
Grace Hopper Conference 2008, Keystone, CO

Introduction to Machine Learning Research on Time Series [1/29/2007]  
Time Series Seminar, Initiative in Innovative Computing  
Harvard University, Cambridge, MA

---

#### AWARDS

---

Robert P. Guertin Graduate Student Leadership Award, 2007

---

#### TEACHING EXPERIENCE

---

Teaching Assistant, Tufts University (Fall 2006)  
Operating Systems, Prof. Alva Couch

---

#### ACTIVITIES

---

President, Tufts Women in Computer Science (WICS) [1/07 – 5/08]  
Member, ACM  
Member, IEEE

---

#### PROFESSIONAL EXPERIENCE

---

**CNET Networks** (now CBS Interactive)

*Software Engineer*, Advertising Technology, CNET Networks [6/00 – 8/04]  
San Francisco, CA & Cambridge, MA (moved to MA in 8/02)  
Promoted from Associate Software Engineer in 6/03

- Contributed to the design and development of the front-end of Madison, CNET’s in-house advertising system, for the data entry of order information into a Sybase database. Framework built in JSP and Java Servlets.
- Designed and developed user interface framework in Java/JSP for the data entry of all online advertisements (i.e. GIF, HTML, Flash, third-party served). Framework empowered end users to uniquely define and deploy templates that specify user interfaces, validation logic and html output for any advertisement type.
- Developed HTML and Excel reports in Java and JSP that report ad traffic and commitment numbers in either HTML or Excel.

- Developed a Java template processing engine using Jakarta's POI software to process Excel spreadsheets coded in a special template language, and output a new Excel spreadsheet with hourly Sales inventory figures. Built administration center for uploading, testing, and deployment of templates.
- Key contributor to dataflow design of advertising creative, from data entry to placement on the web sites. Inherited ownership of the code base for export of ad meta-data and html into XML.
- Technical lead on 4 of 7 rollouts of CNET's advertisement system onto its web properties in 2002. Project managed all engineering tasks.

*Technical Producer*, Ad Development  
CNET Networks, San Francisco, CA

[08/99 – 06/00]

- Expanded CNET's rich media program by writing the first technical specifications for rich media advertisements (Flash, streaming audio/video, etc.), evaluating new advertising technologies and QA'ing rich media advertisements submitted by CNET's premier advertising partners.

*Associate Graphics Producer*, Account Services/Sales  
CNET Networks, San Francisco, CA

[10/98 – 08/99]

- QA'd and inserted all HTML-based advertisements into CNET's ad system with minimal turnaround for CNET's fast-paced, deadline-drive Sales team.