

Vineet Chaoji

CONTACT INFORMATION

Department of Computer Science
Rensselaer Polytechnic Institute
110 8th Street, Amos Eaton 106
Troy, NY 12180 USA

Voice: (617) 784-8700
Fax: (518) 276-4033
E-mail: chaojv AT cs.rpi.edu
WWW: www.cs.rpi.edu/~chaojv

OBJECTIVE

Secure a challenging internship that will further enhance my PhD experience and allow me to explore my research interests in an industry setting

RESEARCH INTERESTS

Algorithms for pattern mining, graph mining and social network analysis, link prediction and path analysis, text mining and applied machine learning

EDUCATION

Rensselaer Polytechnic Institute, Troy, New York USA

Ph.D. Student, Computer Science, September 2004 - Present

- Advisor: Mohammed J. Zaki
- GPA: 3.94

Rochester Institute of Technology, Rochester, New York USA

Master of Science, Computer Science, May 2004

- Thesis Topic: “*Feature Partitioning for the Co-training Setting*”

University of Pune, India

Bachelor of Engineering, Computer Engineering, June 1999

ACADEMIC EXPERIENCE

Rensselaer Polytechnic Institute, Troy, New York USA

Research Assistant

May, 2006 - present

Includes current projects and Ph.D. research.

Center for Discovery Informatics, Rochester Institute of Technology

Research Assistant

December, 2002 - June, 2004

Research projects related to text mining, incremental learning, active learning, and co-training.

Center for Development of Advanced Computing, Pune, India

Research Intern

August 1998 - May, 1999

- Involved in development of a software system (called MANTRA) that translated from English to any Indian language and vice versa.
- Designed and implemented a core component that allowed the retention of formatting attributes of lexical units across translation.
- The whole system was awarded the Computerworld Smithsonian Award (innovations collection).

RESEARCH PROJECTS

- **Orthogonal Graph Mining:** Designed an algorithm for mining a representative set of maximal graph patterns with each member satisfying pairwise orthogonality constraint.
- **Data Mining Template Library:** Involved in the design and development of a generic library for frequent pattern mining. The library allows mining customized patterns. The library can be downloaded from <http://dmtl.sourceforge.net/>.
- **Link Prediction:** involved predicting the likelihood of a link between two entities (in the future) given that they never had a link between them in the past. Explored supervised learning techniques along with Markov chain based techniques to solve the problem. The results were presented at the KDD Challenge.
- **Text Categorization Framework:** Built a framework that enabled use of various information retrieval and machine learning algorithms at different stages of categorization - feature extraction,

cleaning, preprocessing, training and classification. Applied this framework for identifying spam mails.

- **Novelty Detection on Video Streams:** Used low-level features for detecting inconsistent events in a sequence of video data. Applied unsupervised learning to identify inconsistent events along with habituation theory to model the learning aspect.
- **Feature Partitioning for Co-training:** Applied feature splitting to the basic co-training setting to improve performance of co-trained classifiers. Explored the idea of extending co-training to k -training for optimal cost-performance ratio.
- **Authorship Attribution of Text Documents:** Working on developing fault-tolerant sequence mining techniques for capturing stylistic attributes of a text document. The sequences obtained act as features for training a classifier to identify authors.
- **Predicting Protein-Protein Interactions:** Applying link analysis methods along with graph theoretic and unsupervised learning techniques, for the task of predicting interactions between proteins in an interaction network.

PUBLICATIONS

Vineet Chaoji, Mohammad Al Hasan, Saeed Salem and Mohammed Zaki. **ORIGAMI: A Novel and Effective Approach for Mining Representative Orthogonal Graph Patterns.** *Statistical Analysis and Data Mining Journal (under review)*.

Vineet Chaoji, Mohammad Al Hasan, Saeed Salem and Mohammed Zaki. **An Integrated, Generic Approach to Pattern Mining: Data Mining Template Library.** *Data Mining and Knowledge Discovery Journal (under review)*.

Mohammad Al Hasan, Vineet Chaoji, Saeed Salem and Mohammed Zaki. 2006. **Link Prediction using Supervised Learning.** *Workshop on Link Analysis, Counter-terrorism and Security (at SIAM Data Mining Conference)*, Bethesda, MD.

Mohammad Al Hasan, Vineet Chaoji, Saeed Salem, Nagender Parimi and Mohammed Zaki. 2005. **DMTL: A Generic Data Mining Template Library.** *Workshop on Library-Centric Software Design (LCSD), with Object-Oriented Programming, Systems, Languages and Applications (OOP-SLA)*, San Diego.

Mohammed J. Zaki, Nagender Parimi, Nilanjana De, Feng Gao, Benjarath Phoophakdee, Joe Urban, Vineet Chaoji, Mohammad Al Hasan, Saeed Salem. 2005. **Towards Generic Pattern Mining.** *International Conference on Formal Concept Analysis (Invited Paper)*, (LNCS 3403, Springer-Verlag), Lens, France. Shorter version published as invited paper in *Pattern Recognition and Machine Intelligence (PReMI '05)*

Roger S. Gaborski, Vishal S. Vaingankar, Vineet Chaoji, Ankur M. Teredesai. 2004. **VENUS: A System for Novelty Detection in Video Streams with Learning.** *17th International Florida Artificial Intelligence Research Society Conference.*

Roger S. Gaborski, Vishal S. Vaingankar, Vineet Chaoji, Ankur M. Teredesai, Aleksey Tentler. 2004. **Detection of Inconsistent Regions in Video Streams.** *SPIE Proc. 5292, Human Vision and Electronic Imaging IX.*

Vishal S. Vaingankar, Vineet Chaoji, Roger S. Gaborski, Ankur M. Teredesai. 2003. **Cognitively Motivated Habituation for Novelty Detection in Video.** *NIPS Workshop on 'Open Challenges in Cognitive Vision'.*

INDUSTRY EXPERIENCE

Microsoft Corporation, Seattle, Washington USA
Research Intern, Center for Software Excellence

May, 2007 - Aug, 2007

- Built a framework for analyzing failure patterns within a cluster of machines (10K machines) given their monitor (Autopilot) logs.
- Applied itemset and sequence mining techniques to infer probabilistic rules characterizing failure

conditions.

Rogue Wave Software (now acquired by Quovadx), Corvallis, Oregon USA

Software Development Intern

June, 2002 - November, 2002

- Worked on XML Object Link (XOL), which is a part of Rogue Wave's web integration products.
- Designed and developed the module that generates C++/Java code for handling the model group construct of the schema.

Persistent Systems, Pune, India

Software Engineer

July, 1999 - July, 2001

- Member of a team that developed an online collaboration tool for *TeamOn Systems Inc.* (now acquired by *Research In Motion*).
- Involved in initial design and development of the LDAP directory schema for TeamOn.
- Developed a proxy server that allowed users to connect to the mail server using third party clients. The proxy server was designed for scalability and was stress tested.
- Implemented WAP-based access to services such as email, calendar and address book.

PROFESSIONAL
ACTIVITIES

2005-2007: Member of Graduate Admissions Committee, Rensselaer Polytechnic Institute

2004-2005: Member of Graduate Recruiting Committee, Rensselaer Polytechnic Institute

External Reviewer

- 2008: SIGMOD
- 2007: PAKDD, CIKM, ICDM, SBBD
- 2006: ICDM, PAKDD, SIAM SDM, ICDE, COMAD
- 2005: ICDM, KDID, International Conference on Discovery Science, PKDD

COMPUTER SKILLS

Programming Languages : Java, C/C++, Perl, C#, Matlab

Libraries : Jakarta Lucene, WEKA Data Mining Library, BioPerl

Database/Directory : Sybase Adaptive Server, OpenLDAP, Microsoft SQL Server 2000, MySQL

Tools : CVS/RCS, GNU tools (sed, awk, gdb), Apache Ant, L^AT_EX

Web Technologies : HTML, JavaScript, XML/XSLT, XML Schema

Operating Systems : Unix/Linux, Windows, Mac OS, Solaris

REFERENCES

- **Mohammed Zaki**, Ph.D.
Associate Professor, Rensselaer Polytechnic Institute
Email: zaki@cs.rpi.edu
- **Boleslaw Szymanski**, Ph.D.
Professor, Rensselaer Polytechnic Institute
Email: szymansk@cs.rpi.edu
- **David Musser**, Ph.D.
Professor, Rensselaer Polytechnic Institute
Email: musser@cs.rpi.edu