

Apoorv Agarwal

apoorv@cs.columbia.edu
www.cs.columbia.edu/~apoorv
500 Riverside Dr. NY 10027
(917) 545-7835

EDUCATION

Doctor of Philosophy, Computer Science
Columbia University, New York City GPA 4.00/4.00 Jan 2009 - Expected 2012

21st European Summer School in Logic, Language and Information
Bordeaux, France July 20-31, 2009

Master of Science, Computer Science
Columbia University, New York City GPA 3.97/4.00 Sep 2007 - Dec 2008

Bachelor of Technology, Computer Science
Delhi University, New Delhi, India GPA 4.00/4.00 Aug 2002 - Aug 2006

ACADEMIC ACHIEVEMENTS

- Graduate Research Assistantship, Columbia University for Fall '08
- Teaching Assistance Fellowship, Columbia University, Spring '08
- Merit Scholarship and tuition fee waiver, Delhi University (2002-06)

PUBLICATIONS

Apoorv Agarwal and Owen Rambow
“Automatic detection and classification of social events”
In the proceedings of Empirical Methods in Natural Language Processing (EMNLP), 2010, MIT, Massachusetts, USA

Apoorv Agarwal, Owen Rambow and Rebecca Passonneau
“Annotation scheme for social network extraction from text”
In Proceedings of the Fourth Linguistic Annotation Workshop, 2010, ACL, Uppsala, Sweden

Apoorv Agarwal, Fadi Biadisy, Kathleen R. McKeown
“Contextual Phrase-Level Polarity Analysis Using Lexical Affect Scoring and Syntactic N-grams”
In the proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2009), Athens, Greece

Apoorv Agarwal, Owen Rambow, Nandini Bhardwaj
“Predicting Interests of People on Online Social Networks”
In the proceedings of IEEE CSE'09, 12th IEEE International Conference on Computational Science and Engineering, 2009, Vancouver, BC, Canada

RESEARCH INTERESTS AND EXPERTISE

Machine Learning, Natural Language Processing, Sentiment Analysis, Social Media Analysis, Social Network Extraction and Analysis

RESEARCH EXPERIENCE

Social Network Extraction from Texts Aug 2009 - Present

Dr. Owen Rambow, Columbia University

- Extended Automatic Content Extraction (ACE) manual for Entity extraction
- Conceptualized a comprehensive set of social relations for Relation extraction
- Succeeded in finding the most influential characters in the manually extracted social network of *Alice in Wonderland* using network analysis techniques
- Supervising the annotation of ACE 2004 and the Enron E-mail corpus
- Surveyed literature on the application of Tree Kernels for Relation Extraction
- Researching techniques to incorporate deeper semantic analysis of E-mails
- Used Pajek for social network analysis; implementing in Java

Web Document Classification using Tree Kernels Sep - Dec 2009

Class project for Rocco Servedio, Columbia University

- Innovated the application of Tree Kernels with SVM to perform Web Document Classification
- Used BankSearch, a large benchmark data-set, for our experiments
- Demonstrated that the HTML structure of web pages helps predict the domain to which they belong
- Outperformed the baseline system proposed by Sinka and Corne (2002)
- Surveyed literature on String, Spectrum and Suffix Tree convolution kernels
- Implemented in Java and Perl

Predicting Interests of People on Online Social Networks Sep 2008 - Jul 2009

Dr. Owen Rambow, Columbia University

- Introduced a new data set containing both a self-declared friendship network and self-chosen attributes
- Tested the relevance of the friendship network for predicting individual attributes
- Showed that the self-declared friendship network aids in predicting some but not all attributes
- Implemented in Perl and Matlab

Analyzing Sentiment of Opinion-holders in Documents Jan - Sep 2008

Prof. Kathleen McKeown, Columbia University

- Introduced new syntactic and semantic features to predict contextual polarity of phrases in a sentence
- Proposed a novel method of lexical scoring using Dictionary of Affect in Language and WordNet
- Proposed linguistically rich features, based on automated scoring, to capture the affect of context
- Showed a significant improvement over a difficult baseline consisting of lexical N-grams
- Innovated a “norm” that establishes a co-relation between subjectivity and concreteness of words
- Accepted and presented at EACL 2009; implemented in Java.

An Efficient and Scalable Classifier for Data Mining Jun 2005 - Jun 2006

Dr. Satish Chand, Netaji Subhas Institute of Technology, Delhi, India

- Proposed a new algorithm for construction of a decision tree based classifier
- Surveyed literature on association rule mining and classification algorithms
- Implemented C5, ID3 and SLIQ for comparison purposes
- Implemented in C++ and Matlab

PROFESSIONAL EXPERIENCE

Morgan Stanley, New York City

Client Reporting Team

Jul - Aug 2008
(Summer Intern)

- Improved a command-line Client and a SOAP Server for client reporting
- Achieved superior time efficiency and implemented new features
- The project was put into production
- Implemented in Perl

Conexant Systems, India
Broadband Media Processing Unit

Aug 2006 - Aug 2007
(Software Engineer)

- Ported a character driver from VxWorks to Nucleus real time operating system (RTOS)
- Porting was done in a way such that the driver could work for both the RTOS by switch of a flag
- Developed a serial UART driver that can work for multiple baud rates in C

Computer Science Corporation, India
Infrastructure Team

Jun 2005 - Jul 2005
(Summer Intern)

- Proposed a new Entity Relationship design for their 'Employee Training' database
- The design was meant to eliminate data redundancy and inconsistency
- Worked in SEQUEL Enterprise Manager

PROFESSIONAL ACTIVITIES

Reviewer: The 2009 IEEE International Conference on Social Computing (SocialCom-09)

Presentation: Annual KDD Conference at the Oak Ridge National Laboratory, Tennessee, 2008

SKILLS

- OS/RTOS: GNU/Linux, Windows 98/2000/XP, VxWorks, Nucleus
- Programming/Scripting Languages: C/C++, Java, Lex & YACC, Matlab, R, ASM
- Libraries/Tools: STL, WEKA, GTK
- DBMS: MS SQL Server, MySQL, MS Access, MS Excel

INTERESTS

Soccer, Swimming, Jazz Dance and Tabla (percussion instrument in Indian Classical music).

REFERENCES (Columbia University, New York City)

Owen Rambow, Research Scientist at the Center for Computational Learning Systems

rambow@ccls.columbia.edu

(PhD advisor)

Kathleen R. McKeown, Professor in the Department of Computer Science

kathy@cs.columbia.edu

(Departmental advisor)

Adam H. Cannon, Assistant Professor in the Department of Computer Science

cannon@cs.columbia.edu