

Prachi Jain

csz148211@cse.iitd.ac.in, Website: www.cse.iitd.ac.in/~prachij

PhD Student

Indian Institute of Technology, Delhi

RESEARCH INTERESTS

Natural Language Understanding: I am keen to build semantics models for web scale text using Natural Language Processing and Deep Learning techniques guided by Linguistic insights. I am presently working on following projects:

Models for Inference in Knowledge Base: We are working to build deep learning models to make high precision inference on a large scale Knowledge Bases. (arxiv/1706.00637)

Building High precision Inference rule corpus: We are working on algorithms which use linguistic insights to build high precision inference rule corpus. (NAACL HLT 2016)

EDUCATION

Ph.D. (Computer Science and Engineering) July 2014 – Present
Area: Natural Language Processing, Machine Learning
Advisor: Prof Mausam, Associate Professor IIT Delhi
Indian Institute of Technology, Delhi
CGPA : 9.0/10

M.Tech. (Computer Science and Engineering) August 2011 - December 2013
Specialization: Information Security
Advisor: Dr Ponnurangam Kumaraguru, Assistant Professor IIIT Delhi
Indraprastha Institute of Information Technology, Delhi
CGPA : 9.75/10

B.Tech (Hons.) (Computer Science and Engineering) July 2007 - August 2011
Maharshi Dayanand University, Haryana
Cumulative Aggregate : 80.1%

EXPERIENCE

Alcatel-Lucent (Bell Labs), Dublin: Research Intern June 2013 – July 2013

- Guide: Dr Alessandra Sala, Research Scientist Alcatel-Lucent Dublin.
- Handling and Analyzing huge Hive datasets from a leading Telecom Provider.

University Of Southern California, Los Angeles: Research Intern May 2012 – July 2012

- Guide: Dr Kristina Lerman, Project Leader ISI USC.

IIT Delhi: Research Associate January 2014 – April 2014

- Guide: Prof Mausam, Associate Professor IIT Delhi.

IIIT Delhi: Research Assistant January 2013 – May 2013

- Guide: Dr Ponnurangam Kumaraguru, Assistant Professor IIIT Delhi.

Teaching Assistant August 2011 – present

- IIT Delhi: Machine Learning, Artificial Intelligence, Learning Probabilistic Graphical Model, Natural Language Processing
- IIIT Delhi: Mobile Computing, Computer Networks, Digital Electronics.

PreCog Group at IIIT Delhi: Student Researcher January 2012 – December 2013

PUBLICATIONS

- Prachi Jain, Shikhar Murty, Mausam, Soumen Chakrabarti. Joint Matrix-Tensor Factorization for Knowledge Base Inference. Under submission. Code

- Prachi Jain, Mausam. Knowledge Guided Linguistic Rewrites for Inference Rule Verification. Proceedings of NAACL-HLT. 2016. Code
- Prachi Jain. Call Me MayBe: Understanding Nature and Risks of Sharing Mobile Numbers on Online Social Networks. IIIT Delhi Master's thesis, 2013.
- Prachi Jain, Paridhi Jain, Ponnuram Kumaraguru. Call Me MayBe: Understanding Nature and Risks of Sharing Mobile Numbers on Online Social Networks. Proceedings of the first ACM conference on Online social networks. ACM, 2013.(Acceptance rate – 15.9%)
- Kristina Lerman, Prachi Jain, Rumi Ghosh, Jeon-Hyung Kang, Ponnuram Kumaraguru. Limited Attention and Centrality in Social Networks. Social Intelligence and Technology (SOCIETY), 2013 International Conference on. IEEE, 2013.

AWARDS

- Received Microsoft Research travel grant and Xerox Research travel grant for attending NAACL 2016, San Diego.
- Awarded Tata Consultancy Services PhD Fellowship for duration of 4 years from 2016-Present.
- Received Google grant for attending the Lisbon Machine Learning School (LxMLS 2015), Portugal.
- Received *best project award* in Graduate course on Natural Language Processing (NLP) at IIT Delhi. Project name – TweetIE: Extracting Relations from Twitter.
- Selected for the prestigious Viterbi India Research program 2012, to do funded research internship at USC, Los Angeles, CA.
- Full tuition fee waiver for Masters studies at IIIT Delhi (2011)
- Scholarship for securing 3rd position in B.Tech program.

PROJECTS

- Domain Adaptation of word embedding September 2015 - January 2016
 - The vector representation of words is typically estimated using a very large text corpus. This limits its applicability to domains where such a large text corpus is unavailable. In response, we are working on techniques that use small in-domain data for domain adaptation of out-of-domain word vectors.
- TweetIE: Extracting Relations from Twitter September 2014 - December 2014
 - We extracted relation triples (arg1;relation;arg2) from the noisy Twitter data. We first used various tweet normalization techniques and then exploited Brown clustering to refine the tweets and feeded them to OLLIE for extraction. Parallely, we did twitter specific pattern based extractions to identify more triples. We obtained 0.7 precision beating the baseline of 0.6 by SOTA systems on Twitter data.
 - Best project award in graduate level NLP course.
- Error Bounds in Key Expansion of Block Cipher January 2012 - May 2012
 - Analyzed error tolerance in original DES and AES key recovery from corrupt keys, attained by cold boot attack, using block bit error model.
 - Poster accepted at IIIT Delhi Reseach Showcase'12
- Natural Language Question Answering System on Stackoverflow Dataset Using GALAGO January 2012 - May 2012
 - Language: Java

**TECHNICAL
SKILLS**

Languages & Tools: Python, machine learning tools (Theano, scikit-learn), Java, C, C++, Matlab, Git

**RELEVANT
COURSE WORK**

- Natural Language Processing (*IIT Delhi*)
- Advance Topics in Natural Language Processing (*IIT Delhi*)
- Probabilistic Graphical Models (*IIT Delhi*)
- Special Topics in Numerical Algorithm (*IIT Delhi*)
- Machine Learning (*IIT Delhi*)
- Information Retrieval (*IIT Delhi*)