

VIPUL RATHORE

Contact: (+91) 7738259042, rathorevipul28@gmail.com
PhD Student, Dept. of Computer Science and Engineering
Indian Institute of Technology Delhi, New Delhi, 110016, India
Homepage - <http://www.cse.iitd.ac.in/vipulk>

EDUCATION

PhD

July 2019 - Present

Advisor [Mausam](#) (Co-advising with [Parag Singla](#))
Dept. Computer Science and Engineering
College Indian Institute of Technology Delhi
CGPA 8.73/10

MTech.

August 2017 - June 2019

Advisor [Partha Pratim Talukdar](#)
Dept. Computational and Data Sciences
College Indian Institute of Science Bangalore
CGPA 8.5/10

BTech.

July 2012 - May 2016

Dept. Electrical Engineering
College Indian Institute of Technology Bombay
CGPA 7.79/10

PUBLICATIONS

- Vipul Rathore, Aniruddha Deb, Parag Singla, Mausam. “SSP: Self-Supervised Prompting for Cross-Lingual Transfer to Low-Resource Languages using Large Language Models”. (*Under Submission to ACL 2024*).
- Vipul Rathore, Rajdeep Dhingra, Parag Singla, Mausam. “ZGUL: Zero-shot Generalization to Unseen Languages using Multi-source Ensembling of Language Adapters”. EMNLP 2023 (Main Conference). Singapore. December 2023.
- Vipul Rathore, Kartikeya Badola, Parag Singla, Mausam. “PARE: A Simple and Strong Baseline for Monolingual and Multilingual Distantly Supervised Relation Extraction”. Annual Conference of the Association for Computational Linguistics (ACL). Dublin, Ireland. May 2022. [Acceptance Rate: 21%]
- Keshav Kolluru, Samarth Aggarwal, Vipul Rathore, Mausam, Soumen Chakrabarti. “IMOJIE: Iterative Memory-Based Joint Open Information Extraction”. Annual Conference of the Association for Computational Linguistics (ACL). Online. July 2020. [Acceptance Rate: 23%]

WORK EXPERIENCE

Current PhD Thesis project -

Multilingual Information Extraction from Low Resource Languages

June 2021 - Present

- Proposed *ZGUL*, a task-agnostic approach that ensembles Language Adapters in mBERT encoding layers via multi-source training and achieves SOTA results on unseen languages in zero-shot, few-shot and unlabeled data settings. Our technique has been tested for NER and POS tasks on 4 diverse language families (Work accepted in EMNLP 2023 Main Conference).

- Developed *SSP (Self-Supervised Prompting)* framework, which utilizes LLMs such as GPT4 to further improve the results. The idea is to leverage ZGUL or other fine-tuned models to elicit highest confident predictions (based on entropy of predictions), satisfying certain label coverage constraints (for optimal recall), and feed as exemplars for In-Context Learning (ICL) in GPT4.
- Proposed a simple baseline model, *PARE*, for monolingual and multilingual Distantly Supervised Relation Extraction (DSRE). The model achieved state-of-the-art bag-level evaluation results on 4 DSRE datasets (Work accepted in ACL 2022 Main Conference).

2nd year PhD project - *Mitigating gender bias in Machine Learning models* Aug. 2020 - May 2021

- Identified the limitations of existing notions of fairness as they do not take into account the distribution of labels w.r.t. sensitive attributes in the training corpus.
- Proposed our own notion of bias and mitigation technique using constrained learning framework to remove model bias overamplification.

1st year PhD Project - *Open Knowledge Graphs for Multi-Document Summarization* Aug. 2019 - July 2019

- Identified the issue of coverage of relevant entities in summaries of existing state-of-the-art baseline on Multi-News dataset, a large scale dataset for multi-document summarization of news articles.
- Proposed a graph-based approach in which we apply Graph Neural Network (GNN) on top of Open IE based dynamic knowledge graphs to select sentences covering important entities in summarization. Our implemented approach for even extractive summarization beats the previous abstractive state-of-the-art approach.

Teaching Assistant, Indian Institute of Technology Delhi 2019 - Present

Assisted with course material preparation, evaluation and grading for the following courses:

- Natural Language Processing - Responsible for designing and evaluating assignments and exams, preparing course material etc.
- Artificial Intelligence - Responsible for designing AI board game assignment and evaluating student bots by designing and running 3 tournaments each with different board configurations.
- Data structures and algorithms - Responsible for java interface design and writing driver code and testcase generation scripts for an assignment upon application of BST and Heap in real life.

Masters Project - *Ericsson Research Fellow*, Ericsson Research Bangalore 2018 - 2019

Devised solutions to automating troubleshooting using NLP techniques on Ericsson docs, using ideas from the field of reasoning over procedural documents. Demonstrated the efficacy of Graph Convolution Network (GCN) on top of Semantic Role Labeling (SRL) graph applied over procedural documents for tracking entity state (physical location) at each time step.

Research Engineer, Aruba Networks, HPE August 2016 - July 2017

Responsible for *Next-Gen Mobility Controller* evaluation involving board-bringup of Intel, ARM and MIPS processor, carrying out comparative analysis using performance benchmark tools for CPU and network latency. Subsequent data analysis on these results for decision-making on next-gen controller.

AWARDS/ACHIEVEMENTS

- Recognized as [great reviewer](#) for ARR (ACL Rolling Review) October 2023 cycle.
- All India Rank 567 in IIT-JEE amongst 500K students, 2012.
- **Kishore Vaigyanik Protsahan Yojana (KVPY)** fellowship, 2010-11.
- Rank 4 in KVS Junior Mathematics Olympiad, 2010.

- **Ericsson Research Fellowship** for Masters Project at IISc Bangalore (2018-19)

COURSE PROJECTS

COVID-19 Open Research Dataset Challenge (CORD-19), Information Retrieval Course Project
Prof. Srikanta Bedathur, IIT Delhi

Proposed our own algorithm that achieved a rank in top-20 in the Round 1 leaderboard of [CORD-19](#). Further showed that although BM25 doesn't achieve top-50 but, when used on top of BERT's wordpiece tokenizer, gets into top-50 with significant gains in accuracy.

Adversarial attacks on LDA Topic Modelling, ML Course Project *Prof. Chiranjib Bhattacharyya, IISc Bangalore*

Aim was to minimally manipulate the corpus in a way as to fool the topic distribution generated by Latent Dirichlet Allocation. Posed the problem as a bi-level optimization one and successfully changed the rank of target words for particular topics in the *CONG* corpus.

Multi-class Document Classification, Data Analysis and Visualisation Course Project *Prof. Anirban Chakraborty, IISc Bangalore*

A thorough implementation of a variety of machine learning models and subsequent analysis on accuracy vs training time trade-off on *20 Newsgroup* dataset, having 20,000 documents and 20 classes.

SKILLS

- Proficient in Bash scripting, C++, Python, Pytorch.
- Knowledgeable in linear algebra, probability, convex optimization, machine learning etc.

GRADUATE COURSEWORK

- Advanced Data Structures
- Advanced NLP
- Information Retrieval
- Compiler Design
- Natural Language Processing
- Machine Learning
- Practical Data Science
- Numerical Linear Algebra
- Numerical Methods
- Numerical Optimization

OTHER ACTIVITIES

- Reviewer for JAIR (Journal of AI Research) 2021, ARR (ACL Rolling Review) 2021, ARR 2022, ARR 2023, EACL 2024, NAACL 2024, ACL 2024.
- Best Teaching Assistantship Award for COL 106 - Data Structure and Algorithms (Spring 2020-21)
- Attended ICML 2020, ACL 2022, CoDS-COMAD 2020.
- Attended *Amazon Research Days* at Bangalore, 2019.
- Organizer at IISc AI Day 2019, inauguration event for AI Department at IISc Bangalore.

- Organizer at TechFest 2012 and 2013, IIT Bombay's technological festival.
- Performed script-writing, direction and acting in many theatre-level dramatic plays at IIT Bombay.
- Professionally trained for racquet games such as table tennis, badminton and squash.