Suraj Tripathi

Graduate Research Assistant
Language Technologies Institute, Carnegie Mellon University

surajt@andrew.cmu.edu
Github / LinkedIn / Website / Google Scholar
+1-412-954-8105

Education

Carnegie Mellon University   Pittsburgh, USA
Masters in Language Technologies; Fully-funded research master's in NLP (Advisor: Prof. Teruko Mitamura)   *Aug'21 - Present*
Courses: Advanced NLP, Multimodal ML, Multilingual NLP, Question Answering, Computational Ethics, Art and ML
GPA: 4.14 / 4.33

Indian Institute of Technology   Delhi, India
Masters in Computer Science (Advisor: Prof. Jayadeva)   *Aug'15 - Aug'17*
Courses: Machine Learning, Artificial Intelligence, Probabilistic Graphical Models, Numerical Algorithms
GPA: 8.39 / 10.0

Jamia Millia Islamia   Delhi, India
Bachelors in Computer Engineering   *July'11 - July'15*
Courses: Data Mining, Computer Graphics, Database Systems, Operating Systems, Computer Architecture
GPA: 9.31 / 10.0

Selected Publications

[1] **S Tripathi***, S Bansal*, S Agarwal*, T Mitamura, E Nyberg, **PRO-CS : An Instance-Based Prompt Composition Technique for Code-Switched Tasks**, EMNLP'22 [**pdf**]

[2] S Agarwal, **S Tripathi**, T Mitamura, C Rose, **Zero-shot cross-lingual open domain question answering**, Multilingual Information Access Workshop, NAACL'22 [**pdf]**

[3] **S Tripathi***, S Bansal*, S Agarwal*, S Gururaja*, A S Veerubhotla*, R Dutt, T Mitamura, E Nyberg, **R3 : Refined Retriever-Reader pipeline for Multidoc2dial**, DialDoc Workshop, ACL'22 [**pdf**]

[4] P Yenigalla, A Kumar, **S Tripathi,** C Singh, S Kar, J Vepa, **Speech Emotion Recognition using Spectrogram and Phoneme Embedding,** INTERSPEECH'18 [**pdf**]

Skills and Interests

Skills : Natural Language Processing, Machine Learning, Prompt Engineering, Deep Learning, PyTorch, Python
Interests : Large Language Models (LLMs), Parameter Efficient Learning, Multilingual NLP, Question Answering, Dialogue Systems

Work Experience

Amazon, AWS AI Rekognition   Seattle, USA
Applied Scientist Intern (Advisors: Nikolaos Pappas and Daniele Bonadiman)   *May'23 - Aug'23*
- **Cross-lingual RAG:** Working on query transformation and reranking for improving cross-lingual passage retrieval.

Carnegie Mellon University   Pittsburgh, USA
Graduate Research Assistant (Advisor: Prof. Teruko Mitamura)   *Aug'21 - May'23*
- Working on a DARPA project, KAIROS, to identify patterns in articles, induce schemas, predict missing/future events.
- **Event Grounding:** Investigated schema event grounding using transformer based models across a set of extracted elements from multiple documents given its temporal context and various other attributes like entities, relations, etc.
- **Summarization:** Designed approaches to learn an intermediate plan to ground the generation of abstractive summaries in **transformer based seq2seq models**. Our proposed approach achieves SOTA performance on DailyMail and XSum dataset.

Samsung Research Institute   Bengaluru, India
Senior Software Engineer (Bixby, Voice Intelligence R&D)   *Oct'17 - July'21*
- Trained and deployed **NLU models** for **task-oriented dialogue systems** with SOTA performance.
- Engineered a **low latency and low memory** footprint **speech emotion recognition system**. Our approach achieved SOTA performance on IEMOCAP benchmark with **62%** fewer parameters compared to benchmark systems. Published at INTERSPEECH'18.
- Designed lightweight **intent classification and slot tagging models** to map between human commands to low-level actions.

Research Experience

Carnegie Mellon University   Pittsburgh, USA
Graduate Research Assistant   *Aug'21 - May'23*
- **Machine Translation :** Proposed a training paradigm that makes use of a non-deterministic distribution and assigns probability masses to various candidate translations based on their quality. This paradigm incorporates a **contrastive loss** defined over candidate translations produced by pre-trained translation models. [Prof. Graham Neubig | Fall'22]
- **Visual QA :** Trained a model combining **coarse and fine-level features**, enabling semantic reasoning with dynamic feature selection for question answering. Introduced novel Object Feature Extraction and Scenegraph Masking tasks. Explored **prompt learning** in low and high-resource settings, enhancing our proposed methodology's efficacy. [Prof. LP Morency | Fall'21 | **pdf**]
- **Dialogue Systems :** Proposed an approach that employs sparse representations for passage retrieval, a passage re-ranker, the fusion-in-decoder architecture for generation, and a curriculum learning training paradigm. Our approach shows a 12-points improvement in BLEU score compared to the baseline RAG model. [Prof. Teruko Mitamura and Prof. Eric Nyberg | Fall'21 | **pdf**]

Indian Institute of Technology   Delhi, India
Graduate Research Assistant   *Aug'16 - Jun'17*
- **Efficient Neural Networks :** Introduced a novel loss function to achieve sparsity by minimizing a convex upper bound on the Vapnik-Chervonenkis (VC) dimension. Also, analyzed the effectiveness of our proposed loss function in combination with techniques like quantization and pruning. [Prof. Jayadeva | **pdf**]

Honors and Awards
- 1st on UNSEEN track for MultiDoc2Dial, DialDoc Workshop in ACL (2022)
- 3rd on MIA Shared Task, MIA Workshop in NAACL (2022)

Teaching Assistant
- IIT Delhi: Machine Learning (COL774), Computer Networks (CSL 374), Introduction to Computers and Programming (CSL101)

Leadership and Volunteer Experience
- Graduate student mentor at LTI, CMU (2022-23). Internship mentor for undergraduate interns at Samsung Research (2019)