Sabyasachee Baruah

3710 McClintock Ave, Los Angeles, CA 90089
https://sabyasachee.github.io

Email: sbaruah@usc.edu
Mobile: +1-917-361-1742

Education

- **University of Southern California** — Los Angeles, California
  Ph.D - Computer Science; GPA: 3.95/4 — *July 2018 - Dec 2024 (Expected)*
- **Indian Institute of Technology, Kharagpur** — Kharagpur, India
  Dual Degree (B.Tech + M.Tech) - Computer Science & Engineering; CGPA: 9.38/10 — *July 2012 - April 2018*

Research Interests

My research interests lie in **narrative understanding** and **long context modeling**. I use natural language methods to study how stories are told, why some stories appeal more to us, and aid creators in weaving better stories. I have worked in coreference resolution, information extraction, and affective computing, extending to multimodal content.

Projects

- **Character Attribute Extraction**: Explored different prompting methods on large language models to extract attribute frames of characters from movie screenplays *(ICASSP 2024)*.
- **Multimodal Coreference Resolution**: Developed an annotation software to label face tracks with coreferring person names, and tested it on YouTube-8M English news videos *(Google Student Research Internship 2022)*.
- **Character Coreference Resolution**: Designed annotation guidelines and curated a corpus of movie scripts for coreference resolution of characters, and developed scalable coreference resolution models *(ACL Findings 2021 & 2023)*.
- **Fine-grained Opinion Mining**: Developed an opinion mining pipeline to extract opinion holders, targets, expressions, and sentiment polarity in English news articles, as part of a DARPA project (*docker:sabyasachee13/bri-opinion*).
- **Media Representation of Professions**: Frequency and sentiment analysis of professions in media subtitles, covering more than 136K movies and TV-shows between 1950-2017. We created a searchable taxonomy of professions, and curated a sentiment-labeled subtitle corpus, containing more than 3M professional mentions (*PLOS ONE 2022*).

Work Experience

- **Adobe Big Data Experience Lab** — Bengaluru, India
  Summer Research Intern — *May - July 2015*
  - **Data Journalism**: Quantified sufficiency, relevance, and coverage of time-series data.
  - **Recommendation System**: Designed a ranking and recommendation system of data tables for journalists.
- **University of Southern California** — Los Angeles, USA
  Summer Research Intern — *May - July 2016*
  - **Affect Prediction in Movies**: Developed a regression model for continuous valence and arousal prediction in movies using audio-visual features *(ICASSP 2017)*.
  - **Knowledge Transfer**: Used a model trained on discrete-labeled short videos as the base model for gradient boosting for continuous affect prediction.
- **Google** — Los Angeles, USA
  Student Researcher — *May - Aug 2022*
  - **Multimodal Coreference Resolution**: Multimodal analysis of YouTube-8M English news videos.
  - **Annotation Task Design**: Designed an annotation software to answer questions for face tracks.

Publications

- **Sabyasachee Baruah, Shrikanth Narayanan**: Character Attribute Extraction from Movie Scripts using LLMs, IEEE International Conference on Acoustics, Speech, and Signal Processing 2024
- **Sabyasachee Baruah, Shrikanth Narayanan**: *Character Coreference Resolution in Movie Screenplays*, Findings of the Association for Computational Linguistics: ACL-IJCNLP 2023
- **Sabyasachee Baruah, Krishna Somandepalli, Shrikanth Narayanan**: *Representation of professions in entertainment media: Insights into frequency and sentiment trends through computational text analysis*, PLOS ONE 2022
- **Sabyasachee Baruah, Sandeep Nallan Chakravarthula, Shrikanth Narayanan**: *Annotation and Evaluation of Coreference Resolution in Screenplays*, Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021
- **Sabyasachee Baruah, Rahul Gupta, Shrikanth Narayanan**: *A knowledge transfer and boosting approach to the prediction of affect in movies*, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017

Skills and Coursework

- **Programming Languages**: Python, C++, R
- **Frameworks**: PyTorch, Scikit-Learn, HuggingFace, spaCy, Stanza NLP
- **Courses**: Natural Language Processing, Statistics, Dialogue Systems, Knowledge Graphs, Affective Computing

Achievements

- Mentored research interns and USC undergrads to work on emotion recognition in Reddit posts - summer 2021
- Secured perfect 10/10 GPA in second semester of Undergraduate Study - 2013
- Secured national rank 514 and state rank 1 in IIT-JEE examination - 2012
- 20^{th} position (5.5/8 pts) in Telegraph School Chess Competition - Nov 2005