

Sabyasachee Baruah

3710 McClintock Ave, Los Angeles, CA 90089

[Google Scholar](#)

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 - May 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

Building multimodal models for holistic understanding of media stories and fictional characters with a focus towards coreference resolution and sentiment analysis.

PROJECTS

- **Multimodal Coreference Resolution:** Multimodal analysis of YouTube videos.
- **Coreference Resolution in Movie Screenplays:** Designed annotation guidelines for coreference resolution of fictional characters in screenplays, created a coreference-annotated movie scripts corpus, and scaled coreference models to long documents (*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. Currently, I am working to extend it to social media tweets and multilingual documents ([sabyasachee13/bri-opinion](#)).
- **Professions in Media:** 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*).
- **Exaggeration of Health Research in Print and Social Media (Master's Thesis):** Developed a dependency-tree model to find the relational phrase, linking the independent and dependent variables, in health research news (*EACL 2021*).
- **Distributed Charge Scheduling of Electric Vehicles (Bachelor's Thesis):** Designed a distributed algorithm that runs on multiple charge aggregators to schedule the charging time of electric vehicles.

EXPERIENCE

- **Adobe Big Data Experience Lab** Bangaluru, 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
• *Summer Research Intern* May - July 2016
 - **Affect Prediction in Movies:** Developed 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** Venice, Los Angeles
• *Student Researcher* May - Aug 2022
 - **Multimodal Coreference Resolution:** Multimodal analysis of YouTube videos.
 - **Annotation Task Design:** Designed a face-tracking annotation task to answer track-level questions.

PUBLICATIONS

- **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
- **Justin Olah, Sabyasachee Baruah, Digbalay Bose, Shrikanth Narayanan:** *Cross Domain Emotion Recognition using Few Shot Knowledge Transfer*, arXiv:2110.05021 preprint 2021
- **Jasabanta Patro, Sabyasachee Baruah:** *A Simple Three-Step Approach for the Automatic Detection of Exaggerated Statements in Health Science News*, Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics, 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

ACADEMIC 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