

SOPAN KHOSLA

CARNEGIE MELLON UNIVERSITY, PITTSBURGH, 15213
412-370-3255 ◇ sopank@andrew.cmu.edu ◇ [Github](#) ◇ [LinkedIn](#)

PERSONAL SUMMARY

2+ years of industrial research experience spanning multiple research areas, 9 patents filed in the US and multiple publications in peer-reviewed conferences. Primary interests in problems related to natural language understanding (NLP), information extraction from medical records and conversations, dialog systems, and coreference resolution.

EDUCATION

CARNEGIE MELLON UNIVERSITY, PA, USA August 2019 - Present

- MS, Language Technologies Institute
- GPA: 3.96
- Courses: 10-701 Machine Learning, 11-711 Algorithms for NLP, 11-747 Neural Networks for NLP, 80-816 Causal Learning

INDIAN INSTITUTE OF TECHNOLOGY, ROORKEE, INDIA July 2013 - June 2017

- B.Tech, Computer Science
- GPA: 9.43/10, Department Rank 3
- Director's Gold Medal for Outstanding All-round Achievement
- Courses: Data Structures and Algorithms, Artificial Intelligence, Advanced Graphical Networks, Data Mining

EXPERIENCE

RESEARCH ASSISTANT, *Carnegie Mellon University* August 2019 - Present

Working on information extraction and coreference resolution in medical dialogs, under the guidance of [Prof. Carolyn Rose](#).

- Created a system that extracts medically relevant information from doctor-patient conversations. System leverages a multi-speaker BiLSTM for speaker-aware dialogue-modeling, and UMLS (a medical knowledge graph) for grounding raw text into medical concepts.
- Designed an architecture that uses semantic-type information to improve entity coreference resolution on state-of-the-art academic datasets.
- Spearheading the planning of a shared-task (competition) on anaphora and coreference resolution in dialog.

RESEARCH ENGINEER, *Adobe Research India* July 2017 - July 2019

Worked on Natural Language Processing, Social Media Analytics and Marketing Research related projects.

- Used ML and NLP techniques for affect analysis on user-generated content (4 papers, 5 patents filed).
- Developed a technique to quantify latent customer experience from analytics clickstream data (2 papers, 2 patents filed).

TECHNICAL SKILLS

- Python (including scikit-learn, pandas, numpy, Django etc.), Java, node.js, Javascript, React, PHP (Laravel), CSS
- Experience with ML frameworks like Pytorch, Tensorflow, and Keras.
- Data mining and cleaning, full-stack development, statistical data modeling

SELECTED PUBLICATIONS ([Google Scholar](#))

- **Sopan Khosla***, Rishabh Joshi*, Ritam Dutt*, Alan W. Black, and Yulia Tsvetkov. *LTIatCMU at SemEval-2020 Task 11: Incorporating multi-level features for multi-granular propaganda span identification*. SemEval Workshop at CoLing 2020 (**4th rank in the shared-task**). (* Joint First Authors)
- Kundan Krishna, **Sopan Khosla**, Jeffrey P Bigham, Zachary C Lipton. *Generating SOAP Notes from Doctor-Patient Conversations*. ArXiv Preprint 2020.
- **Sopan Khosla**, Niyati Chhaya, Shivam Jindal, Oindrila Saha, and Milind Srivastava. *Do Events Change Opinions on Social Media? Studying the 2016 US Presidential Debates*. In International Conference on Social Informatics, SocInfo 2019.
- **Sopan Khosla***, Kushal Chawla*, and Niyati Chhaya. *Gated Convolutional Encoder-Decoder for Semi-supervised Affect Prediction*. In Pacific-Asia Conference on Knowledge Discovery and Data Mining, PAKDD 2019. (* Joint First Authors)
- Atanu R. Sinha, Deepali Jain, Nikhil Sheoran, **Sopan Khosla**, and Reshmi Sasidharan. *Surveys without Questions: A Reinforcement Learning Approach*. AAAI 2019.
- **Sopan Khosla**, Niyati Chhaya, and Kushal Chawla. *Aff2Vec: Affect-Enriched Distributional Word Representations*. In Proceedings of the 27th International Conference on Computational Linguistics, CoLing 2018.
- **Sopan Khosla**. *EmotionX-AR: CNN-DCNN autoencoder based emotion classifier*. In Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media, ACL 2018 (**Shared-Task Winner System**).