SOPAN KHOSLA

Carnegie Mellon University, Pittsburgh, 15213
412-370-3255 ⋄ sopank@andrew.cmu.edu ⋄ Website ⋄ Github ⋄ LinkedIn

PERSONAL SUMMARY

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

EDUCATION

Carnegie Mellon University, PA, USA — August 2019 - Present

- Master of Science in Language Technologies, Language Technologies Institute
- GPA: 3.96/4.33
- Courses: 10-701 Machine Learning, 11-711 Algo 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 to extract 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 to use semantic-type information to improve entity coreference resolution on state-of-the-art academic datasets.
- Spearheading the planning of CODI 2021 shared-task on anaphora and coreference resolution in dialog. A collaboration with Prof. Massimo Poesio, Prof. Michael Strube, Prof. Carolyn Rose, and Prof. Vincent Ng.

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).

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**, Shikhar Vashishth, Jill Fain Lehman, and Carolyn Rose. *Improving Detection and Categorization of Task-relevant Utterances through Integration of Discourse Structure and Ontological Knowledge.* EMNLP 2020.
- **Sopan Khosla**, and Carolyn Rose. *Using Type Information to Improve Entity Coreference Resolution.* CODI @ EMNLP 2020.
- **Sopan Khosla***, Rishabh Joshi*, Ritam Dutt*, Alan Black, and Yulia Tsvetkov. *LTIatCMU at SemEval-2020 Task 11: Incorporating multi-level features for multi-granular propaganda span identification.* SemEval @ COLING 2020 (**4^{th} 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***, 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**).

INTERNSHIPS

Research Intern, *Adobe Research India* May 2016 - August 2016

Interned at Big-data Intelligence Lab (BIL) at Bangalore, India and completed two independent projects:

- Designed a method to characterize and score users on a website based on their ad blocking tendency.
- Designed a causal method to calculate effectiveness of different anti ad-blocking strategies on a website (1 paper, 1 patent filed).

Software Engineering Intern, *Google Summer of Code* May 2016 - August 2016

Worked as a Core Development Member of FOSSASIA.

- Created a frontend search application over Loklak API.
- Constructed rules for Susi, an AI based chat-bot built on Loklak.
- Created Wordpress plugins which integrate Loklak support to different websites (Blog).

Research Intern, *Symantec R&D Lab India* May 2015 - July 2015

Interned with Digital Signing team at Bangalore, India.

- Developed a tool with an alternate workflow (alias Hash Signing) to digitally sign large files with sizes > 2GB.
- Integrated the tool with Windows SignTool to increase efficiency as well as security of document signing on cloud by 100-1000X (1 patent filed).

RELEVANT COURSEWORK

Machine Learning	Algorithms for NLP	Neural Networks for NLP
Computational Semantics	Causal Modeling	Artificial Intelligence
Design and Analysis of Algorithms	Data Mining and Warehousing	Object Oriented System Design

LEADERSHIP

Co-organizer — *CODI 2021 Shared-task*
Spearheading the planning of CODI 2021 shared-task on anaphora and coreference resolution in dialog.

Internship Mentor — *Adobe Research*
Mentored multiple students at Adobe Research over their summer research internships.

Mentor — *Google Code-In*
Mentored high-school students to contribute to FOSSASIA projects on github as part of the Google Code-In initiative.

Lecturer — *IIT Roorkee*
Organised multiple lectures on Data Structures, OOD, and Machine Learning for students of IIT Roorkee.

Co-Convener Web — *Cognizance*
Led development of IIT Roorkee's annual tech-fest website and supervised branding and social media advertising.

Secretary — *SDSLabs*
Head of administration and web-development at SDSLabs, a student-run technical group at IIT Roorkee.

AWARDS AND ACHIEVEMENTS

2017	IIT Roorkee Director's Gold Medal for Outstanding All-round Achievement
2016	IIT Roorkee Encore Award for All-round excellence
2016	IIT Roorkee Heritage Award Winner thrice, 2014-2016
2016	Microsoft Build The Shield National Champion
2015	Honda Young Engineer and Scientist Award
2015	Microsoft Code.Fun.Do. National finalist in 2015 and 2016
2014	Invited to Republic Day Parade, as a guest of PM of India
2013	All India Rank-7 in JEE-MAINS
2013	All India Rank-1 in CBSE 12th Board Examination (AISSCE)
2013	All India Rank-9 in National Science Talent Search Examination
2012	Kishore Vaigyanik Protsahan Yojana (KVPY) Scholarship
2009	National Talent Search Examination (NTSE) Scholarship