Shashank Siripragada

Personal Data

Webpage: shashanksiripragada.github.io
Github: github.com/shashanksiripragada
Phone: +91-9700109541
email: siripragadashashank@gmail.com

Work Experience

May 19 - Present — **Research Fellow** at **CVIT, IIIT Hyderabad**

- Supervised by Dr. Vinay P. Namboodiri and Dr. C.V. Jawahar
- Research focused on developing NMT systems and curating Multilingual datasets for Low-resourced Indian Languages.

June 17 - May 19 — **Data Scientist** at **Primera Medical Technologies**, Hyderabad

- As a part of Data Science team, my goal was to improve patient outcomes through predictive modelling using patient EHR.
- Built classification models for early detection and intervention in patients at risk of Clostridium Difficile and requirement of home oxygen.
- Built models to detect the patients at risk of overstay and at a risk of SNF placement to assist hospital staff in management of patient logistics.
- Built comprehensive dashboards using EDI 835&837 data for monitoring Insurance Claims and Denials at enterprise scale.

June 16 - July 16 — **Intern** at **Hyundai Motor India Engineering**, Hyderabad

- Developed an application using OpenCV/C++ and Qt to calculate Aperture Ratio from the image of a speaker grill.

Publications

Jan 2021 — Revisiting Low Resource Status of Indian Languages in Machine Translation
J Philip, **S Siripragada**, Vinay P. Namboodiri, C.V. Jawahar
8th ACM CODS and 26th COMAD (CODS COMAD), India

May 2020 — A Multilingual Parallel Corpora Collection Effort for Indian Languages
S Siripragada, J Philip, Vinay P. Namboodiri, C.V. Jawahar
12th Language Resources and Evaluation Conference (LREC), France

Nov 2019 — CVIT's submissions to WAT-2019
J Philip, **S Siripragada**, U Kumar, Vinay P. Namboodiri, C.V. Jawahar
6th Workshop on Asian Translation (WAT), EMNLP, Hong Kong

Education

2013-2017 — B.Tech in Electronics and Communication Engineering,
International Institute of Information Technology, Hyderabad

2010-2012 — Intermediate(12th), **Narayana Junior College**, Hyderabad

Skills

Languages:	Python, SQL, C++, C, MATLAB
ML/Vision:	PyTorch, OpenCV, Scikit-Learn, Pandas
Tools:	Qlik, Tableau, Qt, AzureML Studio
Web:	Flask, HTML, CSS

Projects

Large Scale Parallel Corpus from The Web

- Developed a flask web application to extract parallel corpus for 11 Indian languages from news sources.
- The app contains efficient translation, retrieval and sentence alignment modules enabling working at scale.
- Demonstrated improvements in corpus size and quality with iterative improvements in machine translation and document retrieval performance.
- The work done as a part of this project was published in LREC 2020 and CODS COMAD 2021.
- The corpus was featured in translation tasks at WMT 2020 and WAT 2020, premier translation forums.

MNMT for Indian Languages

- Involved in developing Multilingual Neural Machine Translation (MNMT) systems for Indian Languages, covering 110 language directions.
- Investigated multilingualism, back-translation and efficient strategies to improve NMT performance in low-resourced Indian languages using Transformer architecture.
- The resulting models yield SOTA or competitive performance in WAT Indic translation task leaderboards.
- The design choices and findings of the project are published in WAT 2019 and CODS COMAD 2021.

Research Paper Miner

- Implemented python tool to extract algorithm names from research documents to help users navigate the bulk of scientific research by specific domains.
- The workflow consists of pdf-to-text conversion, tokenization, named entity recognition (NER) and employs cosine similarity on word2vec vectors to determine relevant algorithm names.

Image Captioning

- Implemented an encoder-decoder framework to generate natural language descriptions given an image.
- The image encoder is a Convolutional Neural Net viz. VGG-16 pretrained on ImageNet. For the decoder Vanilla RNN, GRU and LSTM networks have been implemented in PyTorch from scratch, for a comparative study.

Neural Style Transfer

- Implemented the Neural style transfer algorithm proposed by Gatys et al. to render content of one image in the style of another using Convolutional Neural Networks.
- The target image is generated by formulating Content and Style Loss with the respective images and minimizing it. This project was implemented in PyTorch.

Relevant Courses

Natural Language Processing	Computer Vision	Deep Learning
Data Mining & Warehousing	Statistical Methods in AI	Digital Image Processing
Algorithms & Operating Systems	Speech Signal Processing	Information Theory
Computer Networks	Probability	Linear Algebra

Achievements

Nov 2019	Presented our work on NMT for Indian languages at WAT 2019
Oct 2018	Completed 5 part Deep learning specialization by Coursera
July 2016	Presented work done during Internship at *Hyundai R&D Expo 2016*
July 2014	Prathibha Scholarship by *Government of Andhra Pradesh*
May 2013	Secured 259/360 and a rank of 4600 among 1.1 Million people in JEE Mains 2013