Vidhisha **Balachandran**

Graduate Research Assistant · Carnegie Mellon University

5000 Forbes Avenue, Language Technologies Institute, Carnegie Mellon University

(+1) 412-961-2637 | vbalacha@cs.cmu.edu | vidhishanair.github.io | vidhishanair

Education

Carnegie Mellon University

Pittsburgh, Pennsylvania

Masters in Language Technologies (GPA: 3.81/4.0)

August 2017 - Current

- **Advisors: Prof Jaime Carbonell and Prof William Cohen**
- Recipient of Research Fellowship for the entire course duration (Monthly Stipend + Full Tuition Fee Waiver)
- Courses: Algorithms for NLP 11-711, Introduction to Machine Learning 11-701, Neural Networks for NLP 11-747, Structured Prediction for Language and Other Discrete Data 11-763, Probability and Statistics 36-700, Advanced Multimodal Machine Learning 11-777

PES Institute of Technology

Bangalore, India

Bachelors in Computer Science and Engineering (GPA: 9.6/10)

September 2011 - May 2015

- Within the top 5% of the class
- Recipient of MHRD Scholarship 2011-2015 (Full Tuition Fee Waiver)
- Courses : Algorithms, Data Structures, Operating Systems, Data Mining, Natural Language Processing, Database Management Systems, Big Data

Experience

Language Technologies Institute

Pittsburgh, Pennsylvania

Graduate Research Assistant

August 2017 - Current

- Developing models for entity recognition and entity linking for closed domain data
- Building classifiers to learn and predict cases of fraud in the insurance domain
- Developing machine learning models for fraud detection in low resource domain

Flipkart Pvt Limited

Bangalore, India

Software Development Engineer 2

January 2016 - July 2017

- Implemented a self-learning feedback mechanism to regularly optimize models, to account for changing signal behaviour
- Extracted topics from User Review data using Latent Dirichlet Allocation, and analysed sentiment
- Mentored interns on the NLP projects of text analysis; inferring causes for product returns; and contextual keyword extraction using RAKE and Doc2Vec

Flipkart Pvt Limited

Bangalore, India

Software Development Engineer 1

July 2015 - December 2015

- Developed statistical models to generate quality scores for E-Commerce products and listings
- Incorporated the scores into the product search ranking algorithm to ensure better quality products are ranked higher, significantly reducing the return percentage of products by 60 basis points and increasing customer satisfaction by 0.5 Net Promoter Score
- Developed a plug-and-play platform for quick integration of multiple models using Hadoop, Hive, Spark, Redis and Scala
- Scaled the platform to compute scores for 133M entities and developed a low latency system to publish score updates using a Redis based priority queue system and Kafka based messaging system

EMC Corporation India

Bangalore, India

Intern

June 2014 - July 2014

- **Mentor : Dr. Niranjan Thirumale, CTO EMC India**
- Developed a clinical decision system for the EMC LifeCare Platform to assist primary caregiver in disease diagnosis
- Evaluated OpenCDS as a viable clinical decision system and deployed it using decision rules for basic disease profiles

Center for Cloud Computing and Big Data (CCBD), PESIT

Bangalore, India

Intern

June 2013 - June 2014

- **Mentor : Dr. Dinkar Sitaram, Head CCBD**
- Developed hierarchical cross-cloud federation of OpenStack with Amazon EC2 and with other OpenStack cloud deployments
- Designed a project to improve performance of hybrid clouds with federated storage by increasing speed of remote data access and employing intelligent scheduling features

Publications

2018 **Workshop Paper**, Balachandran V, Rajagopal D, Catherine R, Cohen W. Learning to Define Terms in the Software Domain. 4th Workshop on Noisy User-Generated Data, EMNLP. *Brussels, Belgium*

2015 **Conference Paper**, Sitaram D, Phalachandra HL, Harwalkar S, Murugesan S, Sudheendra P, Ananth R, Balachandran V, Kanji AH, Bhat SC, Kruti B. Simple Cloud Federation. Modelling Symposium (AMS) IEEE. *Taipei, Taiwan*

Projects

Latent Structured Representations for Abstractive Summarization

Advisor : Dr Yulia Tsvetkov

TECHNOLOGIES USED: PYTHON, PYTORCH

May 2018 - Present

- Developing methods for latent document representation with focus on structure
- Evaluating latent representations using abstractive summarization with ROUGE metric
- Extracting document level graphs from intermediate latent structure

Self Supervision for Image Captioning using Image Jigsaws

Advisor : Dr Louis-Philippe Morency

TECHNOLOGIES USED: PYTHON, PYTORCH

Sept 2018 - Present

- Exploring usage of self-supervision from solving a Jigsaw puzzle for Multi-modal tasks
- Introducing the Jigsaw puzzle solving task as an auxiliary task for Image Captioning
- Exploring Joint Learning vs Transfer Learning settings to leverage information learned from solving Jigsaw puzzles

Learning to Define Terms in Software Domain

Advisor : Prof William Cohen

TECHNOLOGIES USED: PYTHON, PYTORCH

Aug 2017 - May 2018

- Developed language models to learn to generate definitions for entities in software domain
- Leveraged entity-entity co-occurrence and background ontology for improving the generated definitions by 2 BLEU
- Built a dataset of software entity definitions from Stack Overflow

Table to Text Generation

Advisor : Dr Graham Neubig

TECHNOLOGIES USED: PYTHON, PYTORCH

Jan 2018 - May 2018

- Built a Seq2Seq model for generating biographies of people from Wikipedia Biography Tables
- Used alignments between table and text phrases to improve biographies
- Results were on par with previous State of Art models

Technical Skills

Programming Languages Python, Java
Deep Learning Libraries PyTorch, DyNet, Tensorflow
Database Technologies MySQL, Redis, Hive
Big Data Technologies Hadoop, Hive, Spark
Web Technologies NodeJS, Python Flask, Django

Honours & Awards

2014 **Scholarship Recipeint**, Google Anita Borg Memorial Award Asia Pacific *Tokyo, Japan*
2014 **Scholarship Recipient**, Grace Hopper Conference *Bangalore, India*
2011-15 **First Class Honours with Distinction**, All eight semesters during Bachelors *Bangalore, India*

Leadership

2012-15 **Core Member & Admin**, PES Open Source Community *Bangalore, India*
2013 **Core Organizer**, MIT Media Labs Design Innovation Workshop *Bangalore, India*
2012-14 **Core Organizer**, Incito (Idea hackathon), Hashcode (24-hour hackathon) *Bangalore, India*