SHUBHAM LOHIYA

7119, Hostel 18, Indian Institute of Technology Bombay, Powai, Mumbai - 400076

Research Interests

Knowledge Graphs, Natural Language Processing, Graph Learning, Reinforcement Learning

Education

Indian Institute of Technology Bombay — Mumbai, India

Bachelor of Technology in Mechanical Engineering (with Honors) — *July 2018 – Present*

- Cumulative GPA: **9.56/10.0**
- Minor Degrees: (1) Computer Science and Engineering (2) Machine Intelligence and Data Science

Publications and Working Papers

1. Harkanwar Singh, Soumen Chakrabarti, **Shubham Lohiya**, Prachi Jain, Mausam, **Joint Completion and Alignment of Multilingual Knowledge Graphs**, *Working Paper, to be submitted at KDD 2022*

Research Experience

Joint Completion and Alignment of Multilingual Knowledge Graphs — *Ongoing*

Bachelor's Thesis, Guide: Prof. Soumen Chakrabarti, Department of Computer Science and Engineering — *IIT Bombay*

- **Introduction:** *In the domain of multilingual KGs, Knowledge Graph Completion (KGC) or fact prediction adds valuable information for the Relation Alignment (RA) and Entity Alignment (EA) tasks, and vice versa. The **AlignKGC** framework aims to jointly learn KGC, EA and RA for efficient knowledge transfer across KGs*
- Developed novel datasets from DBP-5L, DBP15k and OpenEA benchmarks for joint analysis of KGC, EA and RA
- Set up **RNM** and **RDGCN** as strong EA baselines and modified them to use mBERT based surface forms
- Discovered **translate+BERT** to be a better backbone than mBERT for KG completion and alignment tasks
- Demonstrated competitiveness of AlignKGC with SOTA approaches while simultaneously learning all three tasks
- Working on a **GNN**-based approach for multilingual KGC which re-formulates the KG as a **tripartite graph**

Entity Alignment in Multilingual Knowledge Graphs — *Jan 2021 - Apr 2021*

RnD Project, Guide: Prof. Soumen Chakrabarti, Department of Computer Science and Engineering — *IIT Bombay*

- **Introduction:** *Entity Alignment (EA) plays a vital role in automatically integrating multiple knowledge bases. Has BERT-INT, which gives almost saturation results on DBP15K, solved the Entity Alignment problem?*
- Conducted detailed EA performance analysis of **mBERT** and **BERT-INT** on WikiData and DBPedia KGs
- Analyzed **entity multiplicity** in the WikiData KG and experimented on novel datasets with higher ambiguity
- Examined the performance of the Interaction module on EA for low-resource language pairs like Bengali-Hindi
- Demonstrated the inadequacy of BERT-INT for EA in presence of low-resource pairs or high entity multiplicity
- Developed a web-application using Django for quick insights and comparison between several EA techniques

Online Reinforcement Learning for Lane Following — *Mar 2021 - Apr 2021*

Guide: Prof. Shivaram Kalyanakrishnan, Department of Computer Science and Engineering — *IIT Bombay*

- **Introduction:** *Lane following is a crucial part of the Autonomous driving problem. Is it possible to learn lane following in an online and efficient fashion using a very low-dimensional discretized state-space?*
- Extracted left & right distance features from dashboard camera feed using Semantic Segmentation and masking
- Employed **Tile Coding** to encode continuous state-variables like velocity, steer and throttle in discrete form
- Handcrafted a multi-objective reward function for the online **Q-learning** agent, which learned a policy on the surprisingly small **5-dimensional** state-space achieving collision-free lane following on the 300m test-track

Professional Experience

Data Scientist | Anheuser-Busch InBev | Growth Analytics Center — *May 2021 – July 2021*

AB-InBev aims to embed Analytics in Business Operations to streamline and optimise tasks

- Engineered a Machine Learning Framework to forecast Accounts receivables for the Mexico region, achieving a mean dispersion of $<$ **5 %** over **5**+ categories and delivering over **$40 million** in working capital benefits
- Developed trend and seasonal features from historical data, and data from Sales, Collections and the Economy
- Trained ensembles of models like **ARIMAX**, **XGBoost**, **Random Forest**, and **MLP** for accurate forecasting
- De-constructed the ML blackbox to analyze the decision making and feature impact using **Shapley Values**
- Awarded the **Pre-Placement Offer (PPO)** for outstanding work during the internship period

Python Developer | **Avrio Energy** *May 2020 – Sep 2020*

The firm is developing AI and IoT powered technology to improve the energy efficiency of businesses

- Designed Schema, Models, and APIs in **Django** for version 1 of Avrio Energy's Outlet Manager android app
- Worked with raw time series data in **InfluxDB** from **1300**+ appliances to perform feature-extraction for **ML**

Winter Intern | **Unacademy** *Dec 2020*

*Unacademy is an Indian online education technology company with **6**+ million active users*

- Planned course map and created content for a **Data Structures and Algorithms** course for GATE aspirants

Major Achievements

- Two-time awardee of the **Institute Academic Prize** (top **1%**) for outstanding academic performance (2021, 2020)
- Ranked **2nd** in a batch of **166** students in the Mechanical Engineering Department (2021)
- Awarded the prestigious **OPJEMS Scholarship** for displaying academic and leadership excellence (2021)
- **Winner** of the Prospect 100 Global Tech Challenge, a Covid-19 Hackathon judged by **Steve Wozniak**. Brainstormed a solution for a globally feasible, logistical solution for **vaccine** storage and distribution (2020)
- Secured **All India Rank 888** out of 1.2 million candidates in the **IIT-JEE** entrance exam (2018)

Key Technical Projects

Web corpus indexing and compression | *CS635: Information Retrieval and Web Mining* *Fall 2021*

- Developed an **inverted index** for a corpus of 50,000+ web documents by encoding **D-gaps** for each token
- Conducted a comparative analysis of Index Compression methods by encoding the D-gap posting lists using various techniques like **Elias Gamma** Coding, **Golomb** Coding, and **Arithmetic** Coding

Profile Recommendation System for Online Dating | *ME781: Statistical ML and Data Mining* *Fall 2021*

- Devised heuristics to construct **compatibility feature vectors** for pairs of dating profiles using profile content
- Synthesized supervision signals using **Proxy Labeling** for learning models like Logistic Regression, SVM, and MLP
- Generated profile recommendations by ranking pair-wise scores, achieved a relevancy score of **79.94%** on test data

Mastering Atari Games using Deep Reinforcement Learning | *CS419: Introduction to ML* *Spring 2021*

- Trained a Deep Reinforcement Learning agent capable of surpassing human performance on classic Atari games like Pong, Breakout and Boxing using **high-level sensory information** in the form of game screen pixels
- Compared the performance of **off-policy** frameworks like **Deep Q-Network (DQN)** and **Double-DQN**
- Leveraged techniques like frame stacking and **experience replay**, allowing the agent to learn from game memory

Automatic Raga Recognition in Indian Classical Music | *IE643: Deep Learning* *Fall 2020*

- Leveraged **tonic** normalized **pitch-tracked** frequencies of a music sample as features for raga classification
- Trained a model based on **LSTM**s with **attention** on random subsequences from the Carnatic Music Dataset
- Achieved a testset accuracy of **96.67%** with 60% majority voting, and **100%** with 50% majority voting

Shortest Path in a Maze | *CS747: Foundations of Intelligent and Learning Agents* *Fall 2020*

- Modelled given 2D mazes as **MDP**s with appropriate states, actions, rewards and transition probabilities
- Compared Howard's Policy Iteration, Value Iteration and Linear Programming algorithms to find shortest path

Image Quilting for Texture Synthesis and Transfer | *CS663: Digital Image Processing* *Fall 2020*

- Implemented a **patch-based algorithm** to synthesize a texture of any desired size from the given sample
- Used a modified quilting algorithm to transfer any given texture to any target image and obtained good results

Other Projects

- Designed a smart **Modular Vertical Farming** unit for climate-independent agriculture in compact spaces (2021)
- Analyzed data of Indian macro-economic indicators to **forecast Consumer Price Index** (CPI) using ML (2021)
- Trained a deep learning framework to perform **Camouflaged Object Detection** in nature images (2020)
- Developed a web-app to do **facial sentiment recognition** on a live video feed using a **CNN** architecture (2020)
- Created a Star Wars themed shooting game using **PyGame** to render graphics, animation, and sound (2020)
- Built sequence model for **Trigger Word Detection**, trained using synthetically generated audio data (2020)
- Constructed an autonomous Line-Following bot using **Arduino UNO**, IR sensors, and a **PID** controller (2019)

Technical Skills

Programming Languages: C++, Python, R, Go, MATLAB, Bash, SQL
Machine Learning: PyTorch, TensorFlow, Keras, OpenCV, Numpy, Pandas, Seaborn
Web Development: HTML, CSS, JavaScript, Node.js
Software: Docker, LaTeX, Git, PowerBI, AutoCAD, SolidWorks, Arduino IDE

Key Coursework

Computer Science: Data Structures and Algorithms, Design and Analysis of Algorithms, Operating Systems[1], Computer Networks[1], Digital Image Processing, Programming for Data Science, Introduction to Machine Learning, ML for Remote Sensing, Deep Learning, Intelligent and Learning Agents (I & II), Information Retrieval and Web Mining, Learning with Graphs, Organizing Web Information[1] , Statistical ML and Data Mining, Advanced Machine Learning[1]

Robotics: Microprocessors and Automatic Control, Robotics, Kinematics and Dynamics of Machines, Machine Design

Mathematics: Differential Equations, Calculus, Linear Algebra, Numerical Analysis

Certifications: 6.86x - Machine Learning with Python (MIT), 6.431x - Probability: The Science of Uncertainty and Data (MIT), 18.6501x - Fundamentals of Statistics (MIT), Deep Learning Specialization (Deeplearning.ai), Fundamentals of Reinforcement Learning (University of Alberta), Game Theory (University of Tokyo)

MOOCs: CS229 - Machine Learning (Stanford), CS231n - Convolutional Neural Networks for Visual Recognition (Stanford), CS285 - Deep Reinforcement Learning (UC Berkeley)

Key Mentoring and Leadership Roles

Institute Secretary Technical Affairs | **Institute Technical Council, IIT Bombay** *Apr 2020 – Mar 2021*

Head of the Electronics and Robotics Club and part of a 23-member core team catering to 5000+ students

- Elected to lead and manage a team of **15+ members** to organize **20+** events, competitions and hackathons and mentor 1200+ electronics and robotics enthusiasts with an annual budget of over **INR 300,000**
- Coordinated the Institute Technical Summer Project program with **70% y-o-y** increase in completed projects
- Initiated the development of **'ERC Wiki'** - a repository of easily accessible resources for enthusiastic learners

Student Mentor | **Student Mentorship Programme, IIT Bombay** *May 2021 – Present*

Selected based on overall performance in a rigorous process comprising of interviews, SOP and peer reviews

- **Institute Student Mentor:** Guiding a group of **12 freshmen** through their first year at IIT Bombay
- **Department Academic Mentor:** Mentoring a group of **7 sophomores** with their academics and research

Teaching Experience

Teaching Assistantships | **IIT Bombay**

Facilitating smooth course organization, grading papers, mentoring students, conducting tutorials and help sessions

- **CS 419 - Introduction to Machine Learning**, Prof. Abir De, CSE Department *Spring 2022*
- **IE 643 - Deep Learning**, Prof. P. Balamurugan, IEOR Department *Fall 2021*
- **ME 119 - Engineering Drawing**, Prof. Deepak Marla, Department of Mechanical Engineering *Fall 2021*
- **MA 106 - Linear Algebra**, Prof. Sudhir Ghorpade, Department of Mathematics *Spring 2021*
- **MA 108 - Differential Equations**, Prof. Prachi Mahajan, Department of Mathematics *Spring 2021*
- **PH 107 - Quantum Physics**, Prof. Shankaranarayanan S, Department of Physics *Fall 2020*

Python is Cool, Kids | **Student-run Summer Course** *Summer 2021*

- Spearheaded a team of students to conduct a summer course for **Practical Python Programming**, consisting of interactive live lectures and guided projects, with **1000+ enrollments**

Extra Curricular Activities and Other Achievements

Achievements	• Selected among the four **delegates from India** to the 5-day virtual "Humanizing Digital 2021" **AI and Data Science** conference at **Chulalongkorn University, Thailand** • Ranked **4** in India's Best Student Contest 2015 organized by RaoIIT amongst **0.3** million participants • Selected among **top 30** students in a Nationwide Aptitude Test conducted by VNIT, Nagpur
Mentorship	• Mentored **9 freshmen** on a project to create an AI agent for mastering the snake game using **RL** • Guided 6 students with reading projects on **Deep Learning** and **Reinforcement Learning**
Technical	• Led a team of **4** to build a radio-controlled trainer aircraft capable of dropping payloads • Constructed an all-terrain obstacle manoeuvring bot controlled using a mobile application • Participated in the **Boeing Aeromodelling Competition** at Techfest 2019, IIT Bombay • Completed a reading project on the use of Deep Learning in **Computer Vision** under SoS 2019
E-Cell, IIT Bombay	• Led a team of **5** organizers during the **Entrepreneurship Summit 2020**, to successfully execute **20+** talks, interviews, and lectures as a **Corporate Relations Coordinator**

[1] To be taken in Spring 2022