

HARSHVEER SINGH

8427337873 • Ludhiana, Punjab

harshveer321.code@gmail.com • github: <https://www.github.com/IIStringll> • DOB: 25 Feb, 2000

EDUCATION

Thapar Institute of Engineering and Technology(TIET)

July '17 - Present

Bachelor of Engineering (B.E.) in Electronics and Computer Engineering

EXPERIENCE & TECHNICAL PROJECTS

Academic Head, ACM Student Chapter, T.I.E.T.

July '19 - July '20

Planned the academic curriculum : topics, number of classes, mentor assignment.

Mentored the Machine Learning research group

Mentored teams for PECFEST Hackathon '19, SATURNALIA Hackathon '19

Tinotifications

Dec '18

Developed an event management and publicity platform for TIET, (currently in use, exclusive for TIET students), backboneed with a Transformer based attention model (Poly Encoders), to pick next favorable task to operate on, ensuring smooth functioning of a given student organization at TIET.

Hackathons

2017-2018

SATURNALIA Hackathon '17, used bayesian prior to regress rural health service data.

PECFEST Hackathon '17, used VAEs for clustering supply chain records of a given organization. Jenson-Shannon divergence metric performed better for estimating the posterior than basic KLdivergence.

PECFEST Hackathon '18, developed a question answering system based on gpt2, to help under privileged students to learn subjects on their own.

RESEARCH EXPERIENCE

Adversarial training for Facebook's blender bot

June'20 - Present

Aiming at a self play regime for conversational agents, and extending that to competitive conversation where an agent discriminates the output distribution of the other agent against human dialogue distribution.

Poly encoder regime for fine-tuning decoder-only model (GPT-2)

May' 20

Decoder model fine tuned like such, apparently is more robust to inductive bias than encoder model even though encoder reached better recall@k/C score.

Analytical study of success of Batch Norm

Nov'19 - Dec'19

Showed that batch normalization smooths the loss surface and how it brings that effect, through the study of eigenvalues of the hessian of input matrix.

Beta2 variation regime for Adam Optimizer

May '18 - July '18

Developed a novel regime for varying beta2 hyper-parameter of Adam (adaptive momentum), preventing Adam from getting stuck in sub-optimal minima. A similar result was also shown in the parallel but completely independent research [Sashank J. Reddi et al.](#)

SKILLS

Mathematics: Probabilistic modeling, Linear algebra, Statistics, Bayesian Inference, Graph Theory, Hyper parameter optimization, Regularization, feature engineering, Classical Optimization algorithms, Heuristic optimization, Analytical Classification, Regression

Programming: Data structures and algorithm design, Predictive Modeling, Python, PyTorch, NLTK, ScikitLearn, C, C++, HuggingFace, NumPy, TensorFlow, Keras, scikit-learn, pandas, matplotlib.

Web: NodeJS, ReactJS, VanillaJS, MongoDB, Nginx, Multiple Third-party APIs.

Utilities: Git/GitHub, Linux(Kernel Programming), Atom, SublimeText, LaTeXs

PUBLICATIONS

The real reason why BatchNorm works

<https://towardsdatascience.com/why-batchnorm-works-518bb004bc58>

Code for Poly-encoder got featured at the official paperswithcode repo

[paperswithcode.com profile of Poly-encoders\[Humeau et al.\]](#),

Code - [IIStringly/Poly-encoders](#)

TALKS AND PRESENTATIONS

Causality and its importance in variational inference and EM, TIET

Jan '20

Inductive bias in machine learning models, TIET

Oct '19

Effect of constraining the posterior to a Gaussian in VAEs, PECFEST

Nov '17

CURRENT WORK

Evolution of ML algorithms from scratch using only basic mathematical operations as building blocks, applying evolutionary methods to automatically find the code for complete ML algorithms.

Finally aiming at creating such a regime of environment and agents, where agents can learn to code their environment and eventually act as an evolved drop-in replacement of their environment, hence bootstrapping the process. Inspired by neural architecture search and AutoML-Zero by GoogleAI.

Analytical study of GPT3 and other predecessors, to figure out why and how they show signs of reasoning, and prove their reasoning to be not falsifiable.

Why deep attention models achieve better adaptability to open domain tasks.

An opposite perspective to transfer learning.

Removing redundancy in GPT3, eventually making it work with minimal computation.