HARSHVEER SINGH

https://llstringll.github.io/
harshveer321.code@gmail.com • Git: @llStringll

EDUCATION

Thapar Institute of Engineering and Technology (TIET) July'17 - Jul'21
Bachelor of Engineering(B.E.) in Electronics and Computer Engineering

SKILLS

Data structures and algorithms, Theoretical machine learning and deep learning, Optimisation algorithms, Loss Landscape Analysis, NLP, NLI, Conversational Agents, Bayesian Inference, C, C++, Python, PyTorch, NLTK, ScikitLearn, HuggingFace, NumPy, TensorFlow, Keras, Pandas, Matplotlib, Shell scripting, Statistics, Probability, Linear Algebra, VertexAI, AWS Sagemaker, Docker

EXPERIENCE AND TECHNICAL PROJECTS

ChartChatAI - visit here
- An inhouse multi modal large model (finetuned Llama) with basic RAG layer to provide candlestick chart analysis in various formats.
- Production stack - nodejs, css/js, weaviate vector db, pytorch, llammacpp
- Also launched on Product Hunt here

Repodox - visit here
- An attempt at very large scale and efficient RAG pipeline over public, personalised GitHub repos, to create a system to easily fetch and discuss and generate specifics from a code repo
- Production stack - nodejs, css/js, weaviate vector db, pytorch, llammacpp
- Is a work in progress

Data Scientist, Applied AI, Vaultedge Pvt. Ltd. Oct'22 - Present
1. Enhanced production model metrics by ~11%, and without regressions.
2. Implemented model interpretability techniques
3. Developed and integrated CI/CD pipelines for seamless model deployment, reducing the need for manual human intervention and accelerating the model update cycle.
4. End-2-end implemented robust data preprocessing pipelines to clean and improve noisy real-world data for language models in production.
5. Trained, optimized, and deployed multi-lingual language models in production

NLP Engineer, AI Team, Mercer-Mettl Jun'21 - Sep'22
1.Improved inter-sentence and intra-sentence cohesion measuring pipeline precision by 25%, involved feature understanding, reducing the output feature space of BeRT by putting a posterior on latent space
2.Implemented production ready email formality checking pipeline for business environments, involved topic modeling of a raw real-life email set that had noisy lexical structure, through improvised LDA
3. Analyzed on-prod spell-check pipeline and suggested and upgrade with specific fine-tuning, that increased recall and precision of the model by ~13% and ~32%.

NLP Engineer Intern, AI Team, Mercer-Mettl Jan'21 - Jun'21
Worked on improving NLI models, improved cohesion detection accuracy on English text by ~16%

Deep Learning Research Intern, CMS Experiment, CERN Jan'21 - May'21
Worked on building a quasi-linear attention model to isolate 'interesting' events from the background during the collision of protons with low-Z targets.

Research Intern, Deptt. of Mathematics, IIT-Mandi Dec'19 - Jan'20

Provided an analytical study on the success of Batch Normalisation [Blog]

Hackathons 2017 - 2018

SATURNALIA Hackathon '17 ranked 1st
PEC-FEST Hackathon '17 ranked 2nd
IITB-TECHFEST Hackathon'18 ranked3rd

RESEARCH EXPERIENCE

Cross-layer residual connection transformer Oct'20 - Nov'20

Developed a novel architecture, which has a recursively "smooth" loss surface, allowing the possibility of reaching more generalized minima, even in the absence of good parameter initialisation.

Adversarial Training for Facebook's Blender June'20 - Aug'20

Created a self-play regime for conversational agents, and extending that to a 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

Showed that a decoder model fine-tuned like such on language modeling apparently is more robust to inductive bias than encoder model even though encoder reached better recall@k/C score

.

Analytical study of the 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 weight matrix. [Blog]

Beta2 variation regime for Adam Optimizer May'18-Jul'18

Developed a regime for varying beta2 hyper-parameter of Adam, preventing Adam from getting stuck in sub-optimal minima. A similar result was also shown in a subsection of Sashank J. Reddi et al.

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 Gaussian in VAEs, IITB-TECHFEST Nov '17