Pratik Deoolwadikar

Software Engineer & Data Scientist

pratik.deoolwadikar@gmail.com +91-9769118579 github.com/pratikdk

WORK EXPERIENCE

Data Scientist, Netcore

Apr 2019 - Nov 2020

- Developed semantic text annotator utilizing Knowledge graphs(wikidata, conceptnet etc), Language models, Meta searching and Information retrieval techniques to enrich raw text.
- Analyzed performance of Attention based pretrained Encoder-Decoder Transformer language models like BERT, Universal Sentence Encoder etc and few Bidirectional and Markov approaches on various NLU, NLG applications like text generation and chatbot.
- Developed novel Cluster-Parser algorithm to obtain custom clusters from time series data frames, crafted features using big analytical queries on analytical db.
- Refined ETL pipeline for Quinto.ai's live bots, utilizing docker and language models.
- Build system services, routing and API endpoints(flask, bottle) to store and process queries.

Data Engineer and Scientist, NanoPrecise Sci Corp

Aug 2018 - Apr 2019

- Developed health monitoring and fault detection system for industrial assets utilizing Signal Processing, Math modeling and Sequential Machine Learning algorithms.
- Analyzed acoustic data from sensors employing domain based noise filtering to discover failure patterns. Designed data labeling pipelines for online learning.
- Designed bootstrap algorithm for battery life estimation of sensors. Explored algorithm performance with changes to spatial positioning and intermittent historical utilization.
- Designed, built improved data pipelines and backend architecture utilizing Multi-Processing and Virtualization along with AWS S3, Redis, Apache Kafka, Docker etc. frameworks.

PROJECTS

Developed a Python NLP Framework, Transformers Keras Dataloader

https://git.io/Jt639

- Enables generator based real-time data feed to Transformer models for downstream training, unlocking the capacity to handle bigger datasets and larger batch sizes.
- Provides support to utilize GPU and Multi-Processing for input processing and computation.
- Added support for custom layer pooling strategies to generate word/sentence input vectors.

Scratch implementation of Machine Learning algorithms

https://git.io/Jt63r

- Raw python implementions for many foundational type of machine learning algorithms along with their Multi flavoured, multi-variate, multi-nomial flavours and Optimizers, Activations, Initializers, in an effort to understand mathematics & architecture backing these algorithms.

Transliteration using Encoder-Decoder Attention model and R-CNN

- Pytorch implemention of Encoder-Decoder Attention model pipeline to transliterate text from source(Hindi) to target(English) language script.

PyTorch implementations of Deep Learning algorithms

- Implemented various flavours for popular Deep learning algorithms like CNN, RNN, LSTM, Encoder-Decoder, etc in PyTorch, as curriculum of Deep Learning course from IIT Madras.

Slot filling using CRF and BiLSTM

- Entity slot identification in BIO format using Conditional Random Fields as slot filter on top of BiLSTM to adapt slot dynamics of training corpus.

Neural Relation Extraction using pretrained Language model

- Semantic relation extraction of marked entities from documents, utilizing language model for obtaining word/phrase representations and downstream classifier to map entity pair similarity to all possible relations.

Finetuning Transformer Language models

- Finetune pretrained parameters of Transformer Language models for text classification task, written a super-fast solution by employing techniques like gradient accumulation, dynamic padding, smart batching and mixed precision.

Multi-Armed Bandit Problem

- Studied various reinforcement learning approaches of exploration and exploitation to solve K-armed bandit problem.

Human Activity Recognition, LSTM on TensorFlow Android

- Realtime activity prediction from continuous spatial data of Accelerometer on Android, to classify amongst six different human activities.

Audience Segmentation, Graph Neural Network

- Segment audiences by categorizing complex relationships using GNN, trained on engineered features from email corpus.

Semantic Topic Clustering, Universal Sentence Encoder

- Clustering topics using attention based language model, to group based on semantic relationships among subject topics for analysis.

Behavioural Cloning, Convolutional Neural Network

- Used CNN to predict steering angle from augmented first person images of road & scene to drive a car in the simulator, as a part of Udacity Self Driving Car nanodegree.

SKILLS

Programming Languages: Python, Java, C++, JavaScript, C, C#, PHP, MySQL, HTML & CSS

Machine Learning/DL Frameworks: PyTorch, TensorFlow, Keras, CoreNLP, SparkNLP, OpenCV, YOLO, Matlab, MLFlow, Scikit, `Scipy, Numpy, Pandas, HuggingFace Transformers

Web, Mobile & Cloud Frameworks: Android framework, NodeJS, Docker, Apache Spark, Kinesis, Kafka, EC2, Kubernetes, Django, PySpark

Databases/Storage/Graphs: MongoDB, MySQL, MariaDB, PostgreSQL, Vertica DB, Redshift, Wikidata, Conceptnet, DBpedia, AWS S3, Redis

Softwares/Tools: Unity 3D, Unreal Engine, Cinema 4D, Photoshop, Illustrator, Blender 3D

EDUCATION

Bachelor of Engineering, Computer Engineering Aug 2018

D.T.E University of Mumbai, India

Diploma, Mechanical Engineering Jun 2015

Maharashtra State Board of Technical Education, Mumbai, India

CERTIFICATION

Machine Learning Engineer, Udacity Aug 2018

- Hands on projects using Machine Learning, Deep Learning, Reinforcement Learning.
- Elementary projects from NLP, Computer Vision/Image Processing etc.
- Exposure to ML/DL frameworks like Tensorflow, Keras, Pytorch etc.

AWARDS

Smart India Hackathon 2017

Awarded by:

- Ministry of Road Transport and Highways, Government of India.
- Persistent Systems Ltd.

LINKS

Github

https://github.com/pratikdk

Website

https://pratikdk.github.io

LinkedIn

https://linkedin.com/in/pratikdeoolwadikar