Nhat Pham

Education

University of Maryland, Computer Science: Data Science, Bachelor's, 2020-2022. *GPA*: 3.92
University of Washington, Data Science and Statistics, 2018 - 2019. *GPA*: 3.97

Work Experience

Software Engineering Intern at Verta   **Summer-Fall 2020**

1. **Dataset and Metadata Versioning Client Interface: Scala**
   - Implemented a functional interface with git-inspired immutable data types (e.g: Repository, Commit) and operations.
   - Integrated with AWS S3 for artifacts storage.
2. **Model Registry and Deployment: Python, Go, MySQL**
   - Implemented Model Registry client interface for packaging machine learning models and deployment-specific requirements.
   - Implemented Endpoint client interface for specifying canary update, autoscaling resources, and making queries.
   - Built a command-line interface that allows users to automate their model registry and deployment workflow.
   - Maintained and improved the quality of the deployment backend, written in Go and MySQL.
3. **Other Tasks: Jenkins, Docker, Kubernetes**
   - Set up Jenkins test pipeline for Scala client, and configured Docker images and Kubernetes pod for it to run.
   - Wrote end-to-end deployment tests for Python client.

Awards and Competitions

Winner of AIVIVN Sentiment Analysis Competition   **Spring 2019**

- Built a system that predicted accurately the sentiment of **over 90%** of product reviews, with a test **F1 metric of 0.90012**.
- Model: word2vec embedding + ensemble of attentional recurrent neural network and residual convolutional neural network.
- GitHub Repository: `https://github.com/petrpan26/Aivivn_1`

Personal and Research Projects

Neural Network Toolbox (nn-toolbox)   **Summer 2019**

- Implemented deep learning procedures and models in PyTorch with a *composable and modular design*.
- Reduced the time to set up model development and prototyping for projects and competitions **from weeks to 1-2 days**.
- GitHub: `https://github.com/nhatsmrt/nn-toolbox`.

Data Collection for Food Desert Prediction Research Project at the University of Maryland   **Spring 2020**

- Scraped food data with selenium, beautifulsoup4.
- Built a pipeline to query and process geolocation data from Overpass and Nominatim APIs.
- Made API calls and geometric queries **3-6 times faster** with multiprocessing and spatial indexing (geopandas + rtree).

Wikipedia Passage Retriever: Python, transformers, Flask, beautifulsoup4   **Winter 2021**

- Implemented a tool to index and retrieve relevant Wikipedia passages, using a pretrained model from *transformers* package.
- Built a Flask application to work interactively with the tool, and a command-line interface to automate the workflow.
- GitHub Repository: `https://github.com/nhatsmrt/wiki-dpr`

Skills and Technologies

Proficient With:

1. *Java*: Object-oriented programming design patterns, JUnit unit testing, multithreading and concurrency.
2. *Python*:
   - Client and Backend Web Development (flask, django, sqlite3). Unit and property-based testing (pytest, hypothesis).
   - Command-Line Interface Development (argparse, click). Working with REST APIs (requests).
   - Data science and machine learning (numpy, pandas, scikit-learn, pytorch, pytorch-lightning). Web scraping (beautifulsoup4).
3. *Scala*: Functional programming principles and design patterns. Tools and technologies familiar:
   - sbt. ScalaTest (behavior-driven and property-based testing). Apache Spark and MLLib (Big Data).
4. *Software Engineering Skills*: Jira. Versioning control with Git and GitHub.

Familiar With:

1. *Full Stack Web Development*: Server-side with Go. Client-side with HTML, CSS, JavaScript, ReactJS.
2. *Other Technologies*: C and C++. Unix and terminal. AWS S3. OCaml. Haskell. Rust. Ruby. Jenkins. Docker. Kubernetes.
3. *Computer Science*: Algorithms and data structures. Relational database and SQL. Operating systems. Distributed systems.

📱 *(857) 300 8127*   •   ✉ *nphamcs@gmail.com*   •   🌐 *nhatsmrt.github.io/*   •   ⌂ *nhatsmrt*
in *nhat-pham-72624b66*