Experience

Machine Learning Engineer @ Solvemate GmbH

Berlin - 02/2020 to Present

I am leading the project to implement free-text input into the existing decision tree-based chatbot framework, which is the most requested feature by our customers. This microservice can be configured to use different methods of word vectorization and text preprocessing. I wrote a custom solution to benchmark the different models and preprocessors used by this microservice. Additionally, I have integrated this microservice into our Kubernetes architecture. This microservice utilizes a modern Python stack based on fastAPI and Pydantic for serving and a modern testing / linting suite (fully pytest-based testing, isort, black, flake8).

Other projects that I am working on include customer feedback analysis and text anonymization. I wrote a highly configurable research tool based on Hydra, which can use word-vector- as well as transformer-based models for feedback analysis. For text anonymization, we are exploring the use of Microsoft's presidio library. In order to apply this to our use case, I made a large OSS-contribution to the project to completely rewrite the python analyzer engine. The primary feature that I added was the ability to use multiple language models at once, which enabled multilingual capabilities to the library. But I also completely modernized their unit tests, optimized the code, and removed a lot of hard-coded variables.

From a software engineering / dev-ops perspective, I have also become quite proficient with web frameworks (flask and fastAPI) and have learned to use Kubernetes with automated deployment via Helm.

Machine Learning Engineer @ i2x GmbH

Berlin - 04/2019 to 02/2020

I primarily worked on NLP projects at i2x. This included coordinating with a team of over 15 labelers to create datasets from our raw data, creating tooling to manage the data pipeline from raw data to production datasets, and even performing labeling myself (all data was German). For these completed datasets, I wrote tooling to train models and gather metrics with libraries like huggingface transformers (BERT), fasttext, sklearn and others. I integrated pytorch models into our GRPC production, which was previously tensorflow only. Another project that I worked on was a sentence level similarity search using word embeddings and approximate nearest neighbors, tool capable of searching tens of millions of utterances in milliseconds. This project was also dockerized and GRPC-based for easy of deployment.

Although my primarily responsibilities were NLP, I experimented with several newer ASR systems (fairseq, wav2letter, and NVIDIA's NeMo) and even contributed several PRs to NeMo. Regarding general software development, code optimizations was a task that I excelled at specifically, I often built optimized dataloaders that could process data online during training, which sped up training and greatly reduced memory requirements for large datasets. I also believe in contributing back to upstream OSS projects and had PRs accepted at huggingface transformers, facebook's FBGEMM, pytorch, pytorch tutorials, and PyThaiNLP.

Lastly, due to my proficiency with linux, I was the ML team's de facto systems administrator for our in-house multi-GPU dev machines, so I'm proficient with the *nix cli ecosystem.

Machine Learning Fellow @ Fellowship.ai

Berlin - 01/2019 to 04/2019

Once again, I developed an end-to-end solution to do image classification on a fashion dataset. We achieved state of the art results in the task classifying a fashion style from the Fashion14 dataset. In preparation for a weekly reading group, I optimized an implementation of memory networks on the babi question answer dataset. Worked on implementing semi-supervised semantic segmentation utilizing Deep Extreme Cut and DeepLab-v3+.

Lead Developer @ PyTorch Audio

Berlin - 06/2017 to 01/2019

I was the lead developer for PyTorch's official audio loading library, torchaudio, and continue to be involved in the project. I work directly with the PyTorch team at Facebook on this project. I have implemented input error checking testing, audio datasets, variable length input collate functions, audio IO funcationality via SoX (Sound eXchange), and added audio transformations (both implemented directly in PyTorch and using PyTorch's c++ hooks for SoX's effects chain). Additionally, I transferred the main PyTorch project's code flaking and documentation standards to this project. In October 2018, I was invited to the PyTorch developer conference for my contributions to the project. I continue to follow the project and contribute when I can.

Machine Learning Engineer (Student Job) @ YEAY GmbH

Berlin - 01/2018 to 06/2018

My primary role was to use object recognition on videos to identify different classes of clothing. I utilized a Mask R-CNN-based network with custom additions for blur detection to select the clearest frames. Data prep included finding an appropriate dataset, writing scripts to translate between various annotation formats (PASCAL, COCO, YOLO) and add missing fields such as segmentation naively or algorithmically (i.e. GrabCut). Additionally, I dockerized my solution to make it more portable.

Co-Founder and CTO @ Cygnus Association Management

Atlanta (Remote) - 08/2014 to 01/2019

Founded an real estate management company to manage residential and commercial properties in Atlanta, GA. Responsible for the organization's technological solutions, including but not limited to Google Apps infrastructure, company web presence, and payment services. Collaborate with other cofounders regarding all strategic aspects of the company. Work 100% remotely.

Senior Consultant @ Navigant

Atlanta - 09/2006 to 04/2013

Completed re-underwrites of over 750 residential home loans for several of the TBTF (Too Big To Fail) US banks in disputes related to the US housing crisis. Managed and reviewed team members' work product. Developed software to automate repetitive tasks that cut the per-loan prep time by an order of a magnitude. Promoted in 2007 and 2010.

Education

- Master of Science in Economics and Management Science from Humboldt Universität in 2018
- Bachelors of Arts in Economics from Emory University in 2005

Languages

- English (native speaker)
- German (professional working proficiency - C1)

| Braunschweiger Straße 54, 12055 Berlin | +49 152 0476 3223 | david@sologourmand.com |