Joshua L. Shapiro

(202) 351-1972 | joshua.shapiro@me.com | jshapiro.info
New York, NY

Experience

Research Engineer April 2020 - Present
ASAPP New York, NY

- Focusing on a language modeling initiative to generate rich coversational embeddings, decreasing the need for annotated data and increasing performance across a variety of production models.
- Researched novel attention-based RNN architecture for hybrid ASR and applied model to production usecases.
- Implemented quickthought-style RNN training regime that decreased model size while increasing performance across a variety of production classification tasks.
- Collaborating on next generation internal machine learning training framework.

Senior Machine Learning Engineer January 2019 - March 2020
ASAPP New York, NY

- Designed and implemented an entity recognition and slot filling service for dialogue systems in collaboration with Research and Product Engineering. The service relied on custom NER models, 3rd party libraries like Duckling, and heuristic approaches to identify, extract, and normalize both generic and domain-specific entites.
- Productionized research prototype for mid-flow branch classification for dialogue systems. This included designing a new service, creating a model evaluation pipeline, formalizing analytics events, gathering annotated data, and refactoring research code.
- Collaborated with deployment managers to standardize and document the process for client update requests for all machine learning services.

Machine Learning Engineer July 2018 - December 2018
ASAPP New York, NY

- Implemented a new service to perform conversation summarization given a prototype model. This included working with Product Engineering to define service and analytics interfaces, decoupling model implementation from service implementation for rapid experimentation, and working with Data Science to analyze online AB test results.
- Updated heuristics component of intent classification service and refactored service code to decrease deploy time for new clients.

Cognitive Software Engineer September 2017 - June 2018
IBM Research Yorktown Heights, NY

- Worked in the Data Centric Systems Department at the intersection of high performance computing and deep learning.
- Researched novel techniques for highly scalable video action classification that at the time outperformed state-of-the-art models in terms of accuracy and training speed.
- Created temporal state detection and clustering algorithms for molecular dynamics simulations.

Publications

ASAPP-ASR: Multistream CNN and Self-Attentice SRU for SOTA Speech Recognition 2020
J. Pan, J. Shapiro, J. Wohlwend, KJ. Han, T. Lei, T. Ma Interspeech

Video Action Recognition with an Additional End-to-End Trained Temporal Stream 2019
G. Cong, G. Domeniconi, J. Shapiro, CC. Yang, B. Chen IEEE WACV

Accelerating Deep Neural Network Training for Action Recognition on a Cluster of GPUs 2018
G. Cong, G. Domeniconi, J. Shapiro, F. Zhou, B. Chen SBAC-PAD

Education

The George Washington University 2013-2017
Bachelor of Science in Computer Science; GPA: 3.89 Washington, DC

Korea University Spring 2015
Exchange Program Seoul, South Korea

Technical Skills

Deep Learning: PyTorch, TorchScript, natural language processing, RNNs, Transformers, CNNs, Autoencoders, cuda, distributed training

Programming: Python, Jupyter, Java, SQL

Technical Tools: Git, Jira, AWS, Docker, Kubernetes, Agile Programming Methodologies