Cara Van Uden

caravanuden.com

Education

Dartmouth College, Hanover, NH
Sept 2015 - June 2019

- BA in Computer Science & BA in Cognitive Science
- **Thesis**: "Comparing brain-like representations learned by vanilla, residual, and recurrent CNN architectures". Oral defense. Received high honors.
- **GPA**: 3.87/4.00, Phi Beta Kappa
- **Academic excellence citations**: Algorithms in Biomedical Data Science (graduate course), Major Seminar in Cognitive Science, Cognition, Introduction to Computer Science

Papers

- **Van Uden, C. E.***, Nastase, S. A.*, Connolly, A. C., Feilong, M., Hansen, I., Gobbini, M. I., & Haxby, J. V. (2018). "Modeling semantic encoding in a common neural representational space." *Frontiers in Neuroscience.*

Projects

- Google text normalization - seq2seq LSTM encoder/decoder (Python, Keras)
- Spotify playlist recommendation - neural collaborative filtering, clustering (Python, Keras, Tensorflow)
- ResNet50 image classification API (Python, Keras, Redis, Docker)
- Parkinson's, cardiovascular disease diagnosis - SVM, logistic regression, random forest (Python, Scikit-Learn)

Skills

- Python (Tensorflow, PyTorch, Keras)
- PySpark/Spark
- SQL
- Hadoop/Hive
- Docker
- Google Cloud Platform
- R
- MATLAB

Awards

- Wayfair Hackathon Winner (2019)
- Phi Beta Kappa (2019)
- Senior Thesis with High Honors (2019)
- Neukom Award for Outstanding Undergraduate Research in Computational Science (2019)
- Cognitive Science Academic Achievement Award (2019)
- Neukom Research Scholar (2018)
- Rewriting the Code Fellow (2017-2019)
- Presidential Research Scholar (2017)

Experience

Machine Learning Engineer - Wayfair
Aug 2019 - present
Competitive Intelligence Matching

- Building out machine learning pipelines and services (Python, PySpark, SQL, Hive, Kafka, Docker, Kubernetes) deployed at scale (>half a billion data points per pipeline per day). These pipelines preprocess data, extract features from text and imagery, and perform model prediction for match generation. Moving pipelines and services to the cloud (GCP).
- Responsible for technical design, implementation, code reviews, critical ops rotation.

Visual Similarity

- Leveraged image embedding (Siamese CNN), image type prediction (EfficientNet), and object detection (YOLOv3) models to incorporate environmental imagery into the visually similar product recommendation (Python, Faiss, Airflow) pipeline. Reduced product coverage gap by over 50% and decreased recommendation pipeline runtime by 20%.
- Won the 2019 Wayfair Hackathon (~100 teams) with "ClusterBrowse" - implemented efficient large-scale (>1M images) unsupervised dimensionality reduction (PCA) and clustering (Faiss, BIRCH) of product image embeddings, with applications for catalog search and visually similar product cluster recommendations. Hackathon browse experience currently in beta testing.

Research Intern - Haxby Lab, Dartmouth College
Jan 2018 - June 2019

- Used hyperalignment (fMRI data alignment technique) and forward encoding models to predict neural responses to naturalistic video stimuli across people (Python). Our hyperaligned forward encoding models improve spatial specificity and model performance compared to previous single- and between-subject methods (paper).
- Compared the representations learned by different CNN architectures to those of the human brain's ventral visual stream. Found preliminary evidence that recurrent and deep residual CNNs learn more brain-like representations than classic feedforward models, and achieved state-of-the-art neural response prediction performance in late-stage visual areas (PyTorch, Python) (thesis).

Teaching Assistant - Computer Science Department, Dartmouth College
Mar 2016 - Mar 2019

- Was a teaching assistant for the classes Introduction to Computer Science (introduction to programming, core algorithms and data structures) (Python) and Foundations of Applied Computer Science (linear algebra, optimization, data analysis and modeling) (MATLAB).
- Held office hours (2-4 hours/week) and graded assignments, labs, and exams.

Research Intern - Just Lab, Carnegie Mellon University
Summer 2018

- Decoded the content of neural representations of emotion in open-source resting state fMRI data from 330 people (MATLAB).
- Partitioned fMRI data with a change point detection algorithm. For each partition, used logistic regression and random forest models to predict emotion state (MATLAB).
- Found significant differences in duration and frequency of emotion among clinical groups - depression, bipolar, anxiety, schizophrenia - compared to controls.

Research Intern - Hassanpour Lab, Dartmouth College
Jan 2016 - June 2017

- Worked on identifying substance use risk using Instagram social media data using a CNN/LSTM ensemble. We used ResNet18 CNNs to extract image embeddings and word2vec and LSTMs to extract text embeddings (Torch). The ensemble was able to estimate the risk of alcohol abuse with statistical significance.
- Created the word2vec and LSTM pipelines used to extract word and sentence embeddings.

Translational Data Science Intern - Celgene
Summer 2016

- Built an exploratory data analytics and visualization tool (R, SQL) for analyzing gene expression and drug response data.
- Used site-wide by scientists for exploratory target deconvolution and validation in translational drug development for blood and bone marrow cancers.