Cory Pruce

874 Lakewood Drive – Sunnyvale CA 94089

✉ corypruce [at] gmail [dot] com • 🌐 cpruce.github.io • in CoryPruce • 🐙 Cpruce

Education

Carnegie Mellon University   **Pittsburgh, PA - Mountain View, CA**
Master of Science in Information Technology, GPA: 3.34, Boeing Scholar, INI Scholar   *May 2014–December 2015*
Coursework: Machine Learning, Packet Switching, Cloud Computing, Mobile Security, Software Reverse-Engineering, Computer Systems, Cyber Security Seminar, Web/Browser Security, Applied Information Assurance, Information Security, Managerial Economics, Business Management

Pitzer College, Pomona College track   **Claremont, CA**
Bachelor of Arts in Computer Science, GPA: 3.01, Pitzer Grant Scholar   *August 2010–May 2014*
Coursework: Algorithms, Data Structures, Operations Research, Distributed Systems, Computer Networks, Statistics, Number Theory & Cryptography, Computability & Logic, Programming Languages, Discrete Mathematics, Software Development

Skills

Software: Python, Scala, C, Java, x86/x64, Sklearn, Tensorflow/Keras, Pytorch, Flink, Kafka, OpenCV, IPython/Jupyter, LaTeX

Natural Languages: *Chinese - Intermediate*

Experience

Parallel Machines   **Sunnyvale, CA**
Machine Learning Software Engineer   *June 2017–Present*

Juniper Networks   **Sunnyvale, CA**
Software Engineer II   *February 2016–June 2017*
- Identified a multi-million/yr reduction in maintenance costs (without taking into consideration opportunity costs) by automating the triaging process. Designed the data analysis, presentation, ETL, feature engineering, and prediction model pipeline for realizing the potential. Found SVM's with TF-IDF to outperform Logistic Regression, Multinomial Naive Bayes, Chi2, and RandomForests. Out of thousands of labels, category accuracy achieved so far is over 50%, with degrees of separation reduced for the majority of the other categories. Using the Wagner-Fischer dynamic programming algorithm, created a personalized variation of the Damerau-Levenshtein distance for category degrees of separation. Investigated RankSVM to increase recommendation robustness.
- Implemented and tested a redundant, persistent storage of MAC address tables for high-availability, performance-critical core network gear. Allowed for dynamic allocation and a theoretically infinite number of addresses. Imperative, highly-demanded feature from customers.
- Improved the robustness, performance, and logging of a kernel tracing tool. Adapted to support all deployed architectures (x86, amd64, arm, arm64, octeon, mips, powerpc). Added functionality to verify structure integrity and give informative reporting if corrupted.
- Solved several customer-facing and critical problem reports.

Kaggle   **Online**
Competitions Contributor   *January 2017–Present*
- Planet - Understanding the Amazon from Space: Used pretrained Resnet[18, 34, 50], Densenet[121, 169, 201], VGG[16, 19], and Inception[V3, V4] for Pytorch models ended with a 17-class sigmoid layer, in order to predict whether or not a feature (water, agriculture, primary rainforest, etc.) exists in the aerial image. Ensembled by class using simple averaging and weighted averaging. Tried cyclic learning rates and test-time augmentations such as random crops, flips, and rotations. Highest individually: 16th out of 600. Team final: 9th out of 938.
- Two Sigma Financial Modelling Challenge: Added leave-one-out cross-validation to a ridge (L2) regressor in order to mitigate overfitting.
- Google Cloud & Youtube Video Understanding Challenge: Experimenting with Google Cloud and the video-level models using Tensorflow.
- Intel & MobileODT Cervical Cancer Screening: Wrote kernel for image feature selection using OpenCV k-means clustering unsupervised learning, basic and adaptive thresholds with contour visualizations, and hue ranges to create cropping masks for a given cervix image. Found non-aggressive, generalizable solution for black borders and some instruments by using the max perimeter of bounding boxes formed through finding contours.

SLAC National Accelerator Laboratory - Google   **Menlo Park - Mountain View, CA**
Graduate Research Assistant   *August 2015–December 2015*
- Developed a framework and dashboard for Powernet, which brings manageability, scalability, reliability, homogeneity, security, and performance to the electrical grid.
- Implemented a persistent object-based storage schema for homehubs and time-series price/power data.
- Displayed real-time data in set interval D3 charts.
- Created a C++ phase-difference test that checks x-intercepts for an Arduino voltage monitor.

Carnegie Mellon University   **Pittsburgh, PA - Mountain View, CA**
Packet Switching Teacher's Assistant   *August 2015–December 2015*
- Mentored students and graded papers, homeworks, and projects.
- Introduced Piazza Q&A to the course, which was well-received.
- Clarified MPLS/GMPLS, ATM, SONET, Switching Fabrics, Queueing, Scheduling.

Hewlett-Packard Enterprise   **Palo Alto, CA**
Software Engineer Intern   *June 2015–August 2015*
o Led the intern research project to build a resource synchronization application.
o Enhanced and maintained services for the Common Web Platform group, particularly in search.
o Integrated a REST-ful documentation service into different API's.
o Helped stage a data migration and created user stories in a Scrum Agile process.

NYIT School of Engineering & Computing Sciences   **New York, NY**
REU Fellow: Cryptography   *June 2013–August 2013*
o Project specifically targeted cryptography for resource-limited devices.
o Implemented in C++ the setup and verification phases of a resource-conscious cryptographic protocol.
o Used an optimized implementation of operations over the finite field GF(256).

Coursework

Machine Learning:
o Constructed, trained, and tested decision trees, a naive bayesian model, a neural network with back propagation, and a Hidden Markov Model alpha-beta/Viterbi evaluator/decoder for Part-of-Speech tagging.
o Plotted the results of linear regression using matplotlib.
o Introduced to Deep Learning and attended Machine Learning weekly seminars with speakers.

Cloud Computing:
o Designed and implemented a web service that uses REST interface to respond to queries on a large twitter data-set (1 TB) which has been transformed and loaded into a database.
o Hosted the web service on Amazon EC2 instances behind an Elastic Load Balancer.
o Designed an efficient Extract, Transform and Load (ETL) process to convert the given dataset (JSON) to comma-separated values that can be imported into the backend database.
o Ran the ETL using Elastic MapReduce offered by Amazon Web Services.
o Selected MySQL as the backend database after comparing its performance with HBase for the given queries.
o Scaled the MySQL backend using horizontal replication. (The queries were read-only).

Distributed Systems:
o Implemented a Dynamic Dining Philosophers problem that ensures safety and progress properties for a priority graph through cleanliness of shared forks.
o Built a storage system that replicates key-value pairs across "chords", or highly-available graphs which find successors in $O(\log N)$.
o Designed and created a distributed Yahtzee tournament system set up in a hierarchical structure in order to leverage parallelizeable control flows and closely monitor corrupt nodes capable of Byzantine failures.

Operating Systems:
o Designed and tested a user-level thread library, mutexes, condition variables, semaphores, reader-writer locks, page fault & exception handlers, and system call stubs.
o Wrote drivers for the console, keyboard, and timer in addition to the game 2048 that used the API.
o Coded a program that traced the function call flow.

Mobile Security:
o Discovered, suggested patch, and compensated for Moderate-Severity vulnerability CVE-2016-3897.
o Implemented app collusion to circumvent the old Android permissions model and exfiltrate sensitive data to a remote server via the browser, while appearing to do an innocuous function.
o Developed a set of scripts and APIs to extract and categorize Android native libraries as part of collecting information on Android developer toolchains.

Packet Switching:
o Implemented in Java a simulator of ATM, MPLS, and GMPLS flows for arbitrary topologies.
o Demonstrated different methods of queueing and queueing for an arbitrary-sized switch.
o Modelled the FIFO, round-robin, weighted round-robin, and weighted fair scheduling techniques.

Computer Systems:
o Wrote in C implicit, explicit, and segregated free-block list dynamic memory allocators.
o Designed buffer overflow exploits, a cache simulator, a caching web proxy, a tiny shell, and dissembled functions to gain specific values.

Reverse Engineering:
o Discovered and analyzed most components of the bank trojan Neverquest.
o Documented unpacking, obfuscation techniques, and C&C communication.
o Used IDA Pro, OllyDbg, OllyDump plugin, and Import Table Reconstructor.
o IDC scripting and uncovering malware driver/device objects amd IRP hooks in kernel debugging with WinDbg.

Applied Information Assurance:
o Exposed to domains such as network traffic monitoring with NTop, persistent data forensics with SleuthKit/Autopsy, memory forensics with Volatility, exploitation with Metasploit/Arbitrage, and web app vulnerability testing with Samurai.
o Deployed a GRR rapid response client-server relationship and wrote a lab exercise that utilizes flows and hunts.

Software Development:
o Wrote a Sudoku player/checker.
o Designed and implemented adventure game following the MVC paradigm.
o Programmed in Objective-C for iOS.

Cybersecurity Seminar:
o Designed and implemented a modular intrusion detection system for modern vehicles using the OpenXC platform.
o Based main component on Snort's rule-based system, especially content-matching.
o Intended for the addition of statistical approaches in order to minimize false positives and negatives.

Web/Browser Security:
o Developed a cookie manager Chrome extension that allows filtering based on time, domain, access, and deletion.
o Wrote, stored, and reflected XSS as well as SQL injection exploits.