

Cem Koc

San Francisco, CA
<http://cemkoc.me>

cemkoc@berkeley.edu
<https://github.com/cemkoc>

EDUCATION

University of California, Berkeley

B.S. in Electrical Engineering and Computer Sciences

May 12, 2017

M.Eng. in Electrical Engineering and Computer Sciences

(concentration: **Data Science & Systems, NLP**)

May 14, 2021

- **Relevant Coursework:** [Data Science](#), [Deep Learning](#), [Natural Language Processing](#), [Parallel Computing & Distributed Systems](#), [Advanced Linear Algebra](#), [Convex Optimization](#), [Computer Vision](#), Operating Systems, Data Structures, Algorithms, Computer Architecture, Machine Learning, Statistics & Probability Theory Artificial Intelligence

INDUSTRY EXPERIENCE

Apple

July 2017 – Present

Software Engineer

Sunnyvale, CA

- Working in the Maps Data Science & Analytics Platform team. Working on creating data pipelines, parallel batch and stream processing algorithms using Spark to do processing of geospatial (GIS) and textual data at scale working with GPS probe, real-time traffic, routing and search datasets.
- Worked on building the POI (place of interest) popularity pipeline with MapReduce, using a distributed deterministic pushdown automaton (DPDA).
- Built an efficient reverse geocoder library that provides a scalable and fast solution when doing point-in-polygon lookups and more
- Created a graph based framework that made it easier to write efficient Spark jobs for complex data transformations supported in Python, Java and Scala. Framework re-uses high level transformation operators built on top of Spark code and allows data scientists and engineers to create production-ready Spark jobs for a variety of tasks
- Created a visual graph editor for creating data science and ETL pipelines which can visually build a Spark job, launch the job and render results
- **Patent:** Co-inventor on a 2019 US Patent Application filed by Apple in distributed combinatorial optimization.

Apple

June 2016 – September 2016

Software Engineering Intern

Sunnyvale, CA

- Worked in Apple Special Projects Group. Worked on creating a distributed combinatorial optimization algorithm in Java to approximate a solution to an NP-hard routing problem. Built perturbation based randomized algorithms on large sparse graphs in Java and Spark to efficiently distribute and find solutions to the problem.
- Built a map and a simple simulation to visualize the output routes using Java and Python. Used a KD-tree to represent the large graph for efficient computations.
- Created an ETL pipeline in Spark to ingest internal data from datasets such as: GPS probe and images and wrote SparkSQL queries for analysis.

TubeMogul (Adobe Inc.)

June 2015 – September 2015

Software Engineering Intern

Emeryville, CA

- Worked in the Ads Machine Learning team. Developed and integrated a Java framework working with Gradle/Maven build systems to collect and process real-time bidding results of digital ads in batch (Kafka/Hadoop/Spark) and in streams using Apache Flink. Created a Bayesian forecasting tool using Python and Java to be used in real-time forecasting.
- Created an ETL pipeline to process peta-bytes of time-series data using MapReduce/Spark to forecast how likely an add will win in a given online auction. Created an auction simulator in back-end using Java and a front-end dashboard to monitor the results and add them to a database to query using SQL in an OLAP system.

PUBLICATIONS & CONFERENCES **Body Lift and Drag for a Legged Millirobot in Compliant Beam Environment**
Cem Koc*, Can Koc*, Brian Su*, Carlos Casarez, Ron S. Fearing.
Published in ICRA, 2019. arXiv 1904.09101.

Terrain Classification with Force-Torque Sensor Equipped Millirobot
Cem Koc*, Can Koc*, Brian Su*, Carlos Casarez, Ron S. Fearing.
Presented at 2016 Bay Area Robotics Symposium (BARS).

RESEARCH EXPERIENCE **Berkeley Artificial Intelligence Research (BAIR) Lab** May 2015 – June 2017
Research Assistant Berkeley, CA

Worked under guidance of Professor Ronald S. Fearing in Biomimetic Millisystems Lab

Research Projects

- **Terrain Classification with Force-Torque Sensor Equipped Millirobot**
Worked with a 10cm long, hexapedal milli-robot with force-torque sensing shell attached on top. Worked on binary classification of densely cluttered terrain using time series force-torque sensor data. Implemented and trained ensemble learning, gradient boosted decision trees, SVMs and neural nets to create a robust model. Our model can identify dense terrains in noisy environments with 97% accuracy.
- **Body Lift and Drag for a Legged Millirobot in Compliant Beam Environment through Intrinsic Force Sensing**
In this work we utilized a (previously developed) low-cost 6-axis force/torque sensing shell to measure the interaction forces between a hexapedal millirobot and a set of compliant beams which models a densely cluttered environment. Built a batch and online data processing framework to analyze large timeseries dataset of sensor readings. Discovered the drag and lift forces, specific resistance required to pass through dense terrains could be measured. We further discovered that, interestingly, some contact between the robot and the compliant beams can lower specific resistance of locomotion.

PERSONAL PROJECTS **PyTorch CodeBERT**: Implemented [CodeBERT](#) on a custom dataset using PyTorch for code search from natural language and translating code to natural language.
Parallelized Point Cloud Similarity: Implemented a parallelized Chamfer distance algorithm in C++ for Point Cloud similarity using a CPU based algorithm using OpenMP/MPI and a GPU based algorithm (for comparison) which utilizes distributed memory paralelism.
Deep Portfolio: Implemented a financial model-free reinforcement learning framework from Jiang et al. for the portfolio management problem. Trained agent utilized policy gradient methods to trade on cryptocurrencies in real life.
Deep Doom: Implemented Prioritized Experience Replay from Schaul et al. to do fast deep reinforcement learning on Doom game. Uses ViZDoom, Keras, OpenAI Gym.
AnimeFaces: Learn to recognize faces in Anime frames! Fine tuned computer vision models such as GoogleNet, AlexNet and VGGFace. Uses Caffe and Keras.

TECHNICAL SKILLS **Interest Areas**: Machine Learning Systems, NLP, Numerical Linear Algebra, Optimization, Parallel Computing, Streaming & Big Data Frameworks
ML/DS Frameworks (order of recency): PyTorch, Numpy/Scipy, Sklearn, Pandas, SparkML, Keras/Tensorflow, Caffe
Computing Frameworks: Spark/MapReduce, OpenMP, MPI, UPC++, CUDA, Docker/Kubernetes/AWS, Kibana/ElasticSearch
Analytics Frameworks (order of recency): Hive, SparkSQL, Zeppelin, MySQL/PostgreSQL, Spark Streaming, Flink
Programming Languages (order of recency): Python, Java, C++, Scala, Go, C