JONATHAN A. CHEUNG

Los Angeles, CA 90025 • jacheung6@gmail.com • github.com/jacheung • (415) 519-2029

Summary

- **6 years of experience using statistics and machine learning for data-driven discoveries**
- **4 publications demonstrating high productivity, clear communication, and detail-oriented nature**
- **3 data science contracts leveraging analytics to provide business solutions**

Experience

Data Scientist – Demand Forecasting and Rate Design, Southern California Gas Co. Aug '20 – Present

- Improved forecasting accuracy of daily load by 29%, translating to cost-savings of $3-5M for customers and investors, via model feature development and application of XGBoost
- Saved 600+ yearly work hours by productionizing and automating a pipeline for daily data query, transform, and forecast of total gas consumption across service territories using Airflow
- Increased operational efficiency by 128% through improving month-to-month work consistency via optimized clustering of 6M meter inspections using k-dimensional tree
- Standardized quantification of scheduling scenarios by developing and implementing KPIs that measure clustering efficiency of inspections and inspection completion dependent on visits and distance

Doctoral Researcher – Hires Laboratory, University of Southern California Sep '14 – Apr '20

- Produced key findings for 2 major grants, increasing funding from $1M to $4.5M over 4 years, by collaborating with colleagues to author 4 manuscripts in high-impact peer reviewed journals
- Established protocols to answer the question of "How does the brain represent touch?" by crafting data pipelines to collect, synchronize, and package 30 million timepoints of sensor motion and neural recordings
- Generated more accurate predictive behavioral and neural models through creative feature engineering via applying domain expertise, time-series filtering and physics models
- Resolved a decade-long debate regarding touch search strategies using predictive behavioral modeling on 16 uniquely extracted touch features – Cheung et al. 2019
- Discovered a neural representation of touch location and hypothesized a circuit model for this sensorimotor transformation using generalized linear models (GLM) – Cheung et al. 2020
- Identified a novel representation of touched object angles, a basis for shape recognition, in the brain using two-photon calcium imaging and GLMs to characterize encoding from elementary forces – Kim et al. 2020
- Promoted a legacy of collaboration by recruiting, mentoring, and training new hires in the scientific method

Data Science Contracts

- **Hires Laboratory, University of Southern California** May '20 – Dec '20
  - Saved 2000+ work hours by automating touch frame detection using TensorFlow and the pre-trained base model, MobileNetV2, to make fast and accurate image classification
  - Extended accessibility of automated touch detection to other research groups by publishing the trained model with a generalizable data pipeline to an open-source Python package (*whacc*)
- **Sene Studio** Aug '19 – Jan '20
  - Ensured data quality and improved accessibility to data by replacing disorganized spreadsheets with an automated extract, transform, load (ETL) workflow with data warehousing in AWS S3 and postgreSQL
  - Reduced order returns and costs by 38% ($58k), leading to a 30% increase in positive customer review, using a boosted regression model to optimize fit for made-to-measure suits and jeans
- **Structure Research** Aug '19 – May '20
  - Standardized metrics, improved data quality, and saved days of manual calculations for a premier data center research group by building an automated analyses pipeline in Python
  - Improved customer experience and web traffic by 20% through application of an interactive data visualization monthly newsletter

Skills

- Tools: Python | Matlab | SQL | Spark | AWS S3
- Python libraries: pandas | NumPy | scikit-learn | SciPy | TensorFlow | PyTorch | PySpark | matplotlib | plotly
- Supervised learning: linear and logistic regression | generalized linear models (GLM) | gradient boosted machines (GBM) | XGboost | time-series forecasting | convolutional neural networks (CNNs) | SVM |
- Unsupervised learning: clustering (k-means, DBSCAN, OPTICS), dimensionality reduction (PCA, t-SNE)

Publications, Awards, and Achievements

Maire P, **Cheung JA,** Hires SA (*in preparation)* - WhACC: an automated approach using CNNs to identify the precise moment of whisker-object contacts

Kim J, Erskine A, **Cheung JA,** Hires SA (2020) – Behavioral and neural bases of tactile shape discrimination learning in head-fixed mice (*Neuron*)

Cheung JA, Maire P, Kim J, Lee K, Flynn G, Hires SA (2020) – Active touch remaps barrel cortex output from a representation of self-motion to object location. *(PLoS Biology)*

Cheung JA, Maire P, Kim J, Sy J, Hires SA (2019) – The behavioral basis of whisker-guided anteroposterior object localization in head-fixed mice. *(Current Biology)*

First Place Presentation in Systems Neuroscience 2018 (*USC Annual Research Symposium*)

Cheung JA, Hsu T, Liang J, Kanoski S (2015) The role of central melanin concentrating hormone signaling in the higher-order control of food intake – Annual DORI Symposium 2nd Place Award, invited talk.

Stokes JA, **Cheung JA,** Eddinger KA, Corr M, Yaksh TL (2013) - Toll-like receptor signaling adapter proteins govern spread of neuropathic pain and recovery post nerve injury in male mice. (*J Neuroinflammation)*

National Institute of Health Intramural Research Training Award 2013-2014 (*National Institute on Aging)*

Service and Broader Impacts

Social Chair - USC Neuroscience Graduate Forum Aug '15 – Aug '16

- Promoted interdisciplinary collaborations through fostering community across graduate disciplines by working with administration to negotiate funding and organize monthly events and celebrations

Educator – California Science Center Nov '14 – Jul '16

- Increased exposure of neuroscience to the general public, as measured by a 64% increase in attendance (from an average of 70 to 115 visitors per hour), through the development of new interactive presentations

Biological Sciences Senator – University of California, San Diego Jun '12 – Jun '13

- Expanded visibility of student research to the San Diego public by collating student research and restarting the Saltman Quarterly, UCSD's primary publisher of undergraduate findings
- Improved placement of biological sciences students into full-time careers, tracked by a 6% increase (+152 students) from previous years, through collaboration with Dr. Gabriele Wienhausen, Faculty Director of Education, to develop mentorship programs and science career fairs

College Council President – University of California, San Diego Jun '11 – Jun '12

- Successfully negotiated a decrease in fee-hike and increased opening hours of UCSD libraries by holding public meetings, collecting opinions of students I was elected to represent, and negotiating with administration

Education

Ph.D. Neuroscience
University of Southern California
Los Angeles, CA / 2020

B.S. Human Biology
University of California, San Diego
San Diego, CA / 2013