Jiawei Gu

LinkedIn: https://www.linkedin.com/in/jiawei-gu/ | GitHub: https://github.com/jiawku
Cell: (801)-637-4021 | Email: jiawku17@gmail.com | Salt Lake City, UT

Summary

I am bioinformatician with **fours years of experience** in academic and industry. I obtained a **MS in biology** and a **MS in computer science** from University of Texas at Dallas. It gives me advantage in combining biology knowledge with programming skills in data engineering and analysis.

Work Experience

Bioinformatics Analyst, Navcian Genomics, Inc. Salt Lake City, 08/2018-Present

I work corporately with Lab Operation Team, Variant Scientist Team, DevOps Team ,and R&D team in Navcian Genomics. My job mostly consisted of three parts:

- ***Production & Automation***
  - Daily clinical control sequencing quality assurance.
  - Designed and implemented the **Tumor Mutation Burden(TMB)** calculation algorithm within the bioinformatics team.
  - Built an **automated pipeline** for clinical control assurance in **python**. It connected multiple **cloud service platforms**(Philips IntelliSpace Genomics Medicine, SalesForce, SeraCare IQ, Slack, and Amazon AWS) via **API**.
  - Created an **Illumina BaseSpace Native APP** for sequencing quality control using **Docker**.
- ***Data Engineering***
  - Created tools to query and clean data from **cloud service platforms** (Philips IntelliSpace Genomics Medicine, SalesForce, Clarity LIMS) via **API** for ETL and R&D.
  - Designed a Bioinformatics **Variant Database** to store clinical mutations.
- ***Data Analysis***
  - Implemented and run validation workflows for Illumina TruSight™ Tumor 170 Sequencing Assay.
  - Designed statistical tests for lab operation quality metrics.
  - Help with other ad-hoc data analysis requests.

Research Assistant, Biology Department, UTD 01/2014-12/2016

I focused on the research of Next-Generation Sequencing, especially on ChIP-seq, RNA-seq, and ChIA-PET Seq result analysis, and using machine learning tools to analysze and predict from data.

- Machine Learning Prediction for RNA-chromosome interaction
  - Wrote a **web crawler** to collect genomic and epigenomic data from **online resources**
  - Cleaned data and used Random Shuffle to generate more negative data points.
  - Applied **SVM**(support vector machine) to generate a model

Article publication: Shi, X., et al. "SMARCA4/Brg1 coordinates genetic and epigenetic networks underlying Shh-type medulloblastoma development." Oncogene (2016).

Skills

Programming languages

Python, Java, JavaScript, R, SAS, SQL

Machine Learning & Statistical model

Support Vector Machine(SVM), Random Forest, Deep Learning, Hidden Markov Model(HMM), Expectation Maximization(EM)

NGS(Next Generation Sequencing) Analysis

ChIP-Seq Analysis ,RNA-Seq Analysis, Gene Ontology Enrichment Analysis,ChIA-PET Analysis, Motif Analysis

Frameworks & Tools

BigData: Apache Spark, Apache Hadoop
Web: MEAN stack(MongoDB, Express JS, Angular JS, Node.js), Java Spring
Deep Learning: TensorFlow, Keras
Containerization: Docker, Kubernetes

Education

01/2017-05/2018

MS, Computer Science; University of Texas at Dallas(UTD); **3.91/4**

08/2013-12/2016

MS, Bioinformatics; University of Texas at Dallas(UTD); **3.73/4**

Certification

- SAS Certified Base & Advanced Programmer for SAS 9
- edX Verified Certificate for Big Data Analysis with Apache Spark
- Coursera Deep Learning Specialization

Side Projects

Dog Breed Identification, Kaggle.com 05/2018

- A Kaggle machine learning competition Project
- Design and built a **convolutional neural network(CNN)** in **keras** to determine the breed of a dog in an image
- Applied a **integrated model** of Xception and InceptionV3 to extract bottleneck features from image
- Used three fully connected layers with **drop out** and **batch normalization** to get predicted probabilities
- Used **Adam** optimilzation algorithm to train the final model with **cross entropy** as loss function
- Get the final validation **accuracy of 99.76%**, rank 102 out of 1286 teams**