

EDUCATION

- **University of Southern California** Los Angeles, CA
Ph.D. in Neuroscience; Published 2 top-journal articles Jul. 2011 – Aug. 2017
- **Tsinghua University** Beijing, China
Bachelor of Science in Physics; Published 1 top-journal article Jul. 2007 – Jul. 2011

PROGRAMMING SKILLS

- **Languages:** R, Python, Java, C#, C++, SQL, Matlab, Mathematica
- **Technologies:** Git, Shiny, MySQL, AWS, Kinect, CKAN, Heroku

WORK EXPERIENCE

- **Innovation Solutions** Santa Ana, CA
Product Development Engineer Intern Summer 2015 & Spring 2016
 - **Summary:** Worked with software engineer team to develop applications to record, analyze and report human (healthy and post-stroke individuals) movement performance during Kinect video games and exercises.
 - **Data Acquisition and Smoothing:** Implemented an application to acquire Kinect skeleton data, smooth it with a spike removal filter and an Unscented Kalman Filter (Matlab & C#). More than 90% of the noise and spikes are filtered out.
 - **Human Joint Angle Inference:** Implemented an application to calculate human anatomical joint angles from Kinect skeleton data (Matlab & C#), allowing further analysis of human movement patterns in anatomical terms.
 - **Web Application:** Designed and implemented a web application to report full kinematics performance history to users and physical therapists (R, Shiny & SQL).

DATA SCIENCE PROJECTS

- **Customer Churn Prediction in Telecommunication Industry:**
 - **Summary:** Developed algorithms to predict customer churn probability based on labeled data via Python and Spark.
 - **Models:** Trained supervised learning models, including Logistic Regression, Random Forest and KNN with parameter regularization.
 - **Evaluation:** Evaluated model performance via k-fold cross validation and confusion matrix.
 - **Feature selection:** Analyzed feature importance to identify most significant factors.
- **Natural Language Processing Projects:**
 - **Word Prediction:** Developed an app enabling prediction of most likely following words in real time while a user is typing. Tokenized 2GB text data to train a ngram model. Deployed the app via Shiny framework.
 - **Topic Modeling:** Grouped unlabeled textual documents and inferred latent semantic structures via Python. Features extracted by Term Frequency - Inverse Document Frequency (TF-IDF). Trained K-means and Latent Dirichlet Allocation(LDA) models. Identified latent topics and keywords of each document for clustering and calculated document similarity. Visualized results by Principal Component Analysis (PCA).

RESEARCH PROJECTS

- **Human Movement Simulation:** Simulated human arm movement in state space with 2-link arm model, via Optimal Control theory implemented by iterative Linear-Quadratic Regulator and Dynamic Programming in Matlab to investigate how movement accuracy affects velocity.
- **Human Movement Performance Analysis:** (1) Analyzed 5+GB kinematics data from Arneo@Spring in csv files from 50+ users in Matlab; and (2) Organized 5+GB kinematics data from 40+ users in text files format into MySQL database (therefore increased efficiency by 1000 folds) and analyzed in R to investigate how movement patterns correlate with stroke recovery.
- **Human Movement Performance Modeling:** Developed dynamical State Space Model with mixed effects to investigate the effects of rehabilitation training, therefore allowing customization of training schedule and prediction of future performance for each individual.