

EDUCATION

- **University of Southern California** Los Angeles, CA
Ph.D. in Neuroscience Jul. 2011 – Aug. 2017
- **Tsinghua University** Beijing, China
Bachelor of Science in Physics Jul. 2007 – Jul. 2011

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, 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#)
 - **Human Joint Angle Inference:** Implemented an application to calculate human anatomical joint angles from Kinect skeleton data (Matlab & C#)
 - **Web Application:** Designed and implemented a web application to report full kinematics performance history to users and physical therapists (R, Shiny & SQL)

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 Armeo@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 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, to customize training schedule and predict future performance for each individual

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 and Topic Modeling:**
 - **Summary:** Clustered unlabeled textual documents into groups and inferred latent semantic structures via Python
 - **Preprocessing:** Preprocessed text by tokenizing, stemming and stopwords removing, and extracted features by Term Frequency-Inverse Document Frequency (TF-IDF) approach
 - **Models:** Trained unsupervised learning models: K-means Clustering and Latent Dirichlet Allocation (LDA)
 - **Results:** Identified latent topics and keywords of each document for clustering and calculated document similarity
 - **Visualization:** Visualized results by Principal Component Analysis (PCA)

PROGRAMMING SKILLS

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