Chunji Wang

Los Angeles, CA 90025

chunjiw@gmail.com, github.com/chunjiw

(213) 344-7935

EDUCATION

- **University of Southern California** *Ph.D. in Neuroscience* Los Angeles, CA, *2011/07 – 2017/08*
- **Tsinghua University** *B.S. in Physics* Beijing, China, *2007/07 – 2011/07*

SKILLS

- **Theories:** Machine Learning, Natural Language Processing (NLP), Deep Learning, Statistics
- **Languages:** Python, R, Scala, Matlab, SQL, Java, Mathematica, C#
- **Technologies:** Git, Shiny, MySQL, Spark, AWS-EMR, Kinect, CKAN, Linux

WORK EXPERIENCE

- **Rule14 LLC** *Data Scientist* Santa Monica, CA, *2017/10 – present*
  - **Utility Load Forecast:** Built utility load forecast models for millions of customers; integrated it with relational databases; automated back-end model training and testing.
  - **Customer Churn Prediction:** Built Random Forest models to predict churn probabilities of millions of customers via Spark under Amazon EMR framework.

- **Innovation Solutions** *Data Scientist Intern* Santa Ana, CA, *2015/06 – 2016/05*
  - **Summary:** Developed applications to record, analyze and report human movement kinematics and performance during Kinect video games and exercises.
  - **Data Acquisition:** Implemented a body recorder application to acquire Kinect skeleton data and smooth it with an Unscented Kalman Filter (Matlab & C#), which reduced signal noise by $\sim 90\%$.
  - **Data Analysis:** 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.
  - **Report and Visualization:** Designed and implemented a web application to report full kinematics performance history to users and physical therapists (R, Shiny & SQL) with intuitive graphics.

DATA SCIENCE PROJECTS

- **Word Prediction:** Developed the application wordpred to predict the most likely following word in real time while a user is typing. Tokenized 2GB text data to train an ngram model. Deployed via the Shiny framework.

- **Topic Modeling:** Grouped unlabeled textual documents and inferred latent semantic structures via Python. Extracted features 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

- **Movement Planning:** Simulated human arm movement in state space via Optimal Control Theory implemented by iterative Linear-Quadratic Regulator and Dynamic Programming in Matlab. Showed that a moderate movement velocity achieves better accuracy than a slow or fast velocity.

- **Movement Learning:** Analyzed 10+GB upper extremity kinematics data via Matlab, R, and SQL. Extracted movement patterns and variability patterns via dimension reduction methods. Showed that healthy movement patterns facilitate movement learning and recovery.

- **Movement Learning Prediction:** Developed dynamical State Space Model with Mixed Effects to investigate the effects of rehabilitation training, allowing customization of training schedule and prediction of future performance for each individual.

HOBBIES

Reading, playing badminton, listening to Broadway musicals, recording audiobooks, watching standup comedies