Chunji Wang

chunjiw@gmail.com, github.com/chunjiw

Los Angeles, CA 90025
(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, SQL, Java, C++
- **Technologies:** MySQL, Spark, AWS-EMR, Tesseract OCR

Work Experience

- **Rule14 LLC** *Data Scientist* — Santa Monica, CA, *2017/10 – present*
  - **Utility Load Forecast:** Built utility load forecast models for millions of customers. Adopted robust and reliable linear models with R that have not failed so far. Integrated them with MySQL databases. Automated back-end model training and testing. Achieved a low Mean Absolute Percentage Error (MAPE) of 4%.
  - **Customer Churn Prediction:** Built Gradient Boosted Trees via xgboost with R to predict churn probabilities of millions of customers. Also attempted Survival Models with R and Random Forest Models with Spark.
  - **Optical Character Recognition:** Developed a software with python that extracts field values from scanned documents and generates spreadsheet reports. Adopted Tesseract OCR Engine and OpenCV library. Achieved perfect accuracy.
  - **Document Classification, NLP:** Built multiclass classification models with python to classify legal documents (about 40 classes). Extracted features with Bag of Words and TF-IDF techniques. Attempted many models including Random Forest, SVC, Naive Bayes, and Deep Learning. Achieved 80% accuracy.
- **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.

Other Projects

- **Autocomplete Input Method:** Developed an application wordpred to predict the most likely following word in real time while a user is typing. Tokenized 2GB text data to train a bigram model. Deployed via Shiny framework.
- **Topic Modeling, NLP:** 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).
- **Customized Vocabulary Review:** Developed an vocabulary review app that extracts new words from the user's Google Search history and generates vocabulary cards in html automatically for later review. Obtained word definitions via Oxford Dictionary API.

Publications

- **Wang C** et al. Spin-orbit coupled spinor Bose-Einstein condensates. *Physical Review Letters.* 105, 160403
- **Wang C** et al. The duration of reaching movement is longer than predicted by minimum variance. *Journal of Neurophysiology.* Vol.116 no.5, 2342-2345
- Martinez C, **Wang C**. Structural constraints on learning in the Neural Network. *Journal of Neurophysiology.* Vol.114 no.5, 2555-2557