Sen (Forrest) Yang

96 Frelinghuysen Road, Piscataway, NJ | +1 (848)-565-5991 | sy358@rutgers.edu

LinkedIn | Website | Google Scholar

Objective

Fresh Ph.D. graduate seeking a full-time position as a data scientist or machine learning scientist. Experience includes Data Mining, Machine Learning, Deep Learning, Algorithms, Process Mining, Software Engineering in Data Visual Analytics.

Technical Skills

Programming:	Java, Python (sklearn, pandas, TensorFlow, Keras), MATLAB, R, C++, (JS, HTML, CSS), Android
Data Mining:	Modeling, Visualization, Process Mining, Recommender System, NLP
Machine Learning:	Supervised Learning, Unsupervised Learning (clustering, anomaly detection), Feature Engineering
Deep Learning:	RNN, CNN, Autoencoder
Statistics:	Probability, Distribution, Statistical Inference, Hypothesis Testing, Bayes Theorem, ANOVA
Data Visualization:	Python (matplotlib, seaborn), JavaScript (D3.js), Java Swing, MATLAB, R, Weka, Tableau
Big Data and Database:	MySQL, Mongo DB

Education

Rutgers University, the State University of New Jersey — Piscataway, NJ
- *M.S & Ph.D.*, in Electrical & Computer Engineering, GPA 3.9/4.0 — Sept 2013 – Expected Oct. 2018

Nanjing University of Posts and Telecommunications, China — Nanjing, China
- *B.A.*, in Communication Engineering, GPA 84/100 — Sept 2008 – June 2012

Working Experience

Nokia Bell Labs — Murray Hill, NJ
- Data Scientist Intern (& Co-op), Data Science Group — May 2017 – Feb 2018
- Participated Project: Automated Machine Learning
- Selected to present to Marcus Weldon, president of Bell Labs and CTO of Nokia.

Rutgers University & Children's National Medical Center (CNMC) — Piscataway, NJ
- Research Assistant & Data Analyst — Aug 2014 - Present
- NIH project: Automatic Workflow Capture & Analysis for Improving Trauma Resuscitation Outcomes

Huawei Technologies Co. Ltd. — Shenzhen, China
- Technical Engineer on GSM, LTE network maintenance — Aug 2012 – June 2013

Research and Projects

1. Automated Feature Extraction from Time Series with Deep Learning — *Bell Labs*, 2017 - Present
- Research the use of Autoencoder (with LSTM, Bi-LSTM, CNN, etc.) for unsupervised feature learning for time series
- Compare the feature learning performance of deep learning approaches with traditional statistical approaches (TSFresh)
- Evaluate in both supervised and unsupervised ways with UCR data collections and two other datasets collected at Bell labs

2. Deep Learning Framework for Next Medical Treatment Activity Recommendation — *NIH Project*, 2017 - 2018
- Developed a context-aware RNN based recommender system to provide runtime treatment recommendations
- Invented Act2vec, a method to embed human activities or events into numerical vectors via a neural net
- Proposed a novel data augmentation algorithm that can fabricate synthetic patient data that closely resembles authentic data
- Evaluated our system on two medical processes and achieved a top-1 accuracy of 0.46 and a top-3 accuracy of 0.77 (from 15 classes)

3. Nature Language Processing (NLP) for Patient Injury Prediction — *NIH Project*, 2018 - Present
- Built a system to predict the likelihood of patient injuries according to the "mechanism of injury" description in the EMS report
- Used an attention-based Bidirectional recurrent neural networks with pre-trained word vectors (Glove, Stanford NLP)
- Compared to traditional NLP methods (bag-of-words, n-gram with simple classifiers)

4. Smart Trauma Resuscitation Decision Support System — *NIH Project*, 2014 - Present
- Design a computerized decision support system to monitor trauma resuscitation workflows and alter medical team of errors
- Develop a sensor-based (RFID, Video, and microphone) system to automatically identify medical team activities

- Analyze medical team activities, discovering treatment patterns, detecting human errors and extracting medical knowledge

5. Prototype-based Recommender System for Medical Treatment Procedures *NIH Project*, 2015 - 2017
- Presented a data-driven recommender system that provides step-by-step treatment recommendations
- Developed a time-warping-based temporal sequence similarity measure (TwS-PT) and a method for calculating prototype sequence
- Tested our methods on three real-life medical processes and achieved accuracy up to an F_1 score of 0.77 (0.37 for baseline)
- Implemented as a web app (VIT-PLA 2.0) using D3.js, JSP, Java and includes interactive visual functions

6. NBA Game Winner Prediction (Python, MySQL) *Course Project, Rutgers*, 2014
- Crawled 1229 games (and 61 attributes for each game) from ESPN website for NBA 2013 regular season
- Cleaned (e.g., "New Orleans Hornets" → "New Orleans Pelicans") and transformed (moving average of last 10 games) the raw data
- Achieved the best accuracy (69%) with Support Vector Machine (SVM), compared to baseline ZeroR (61%).

7. Health Monitoring Analytics based on Twitter (Java, MongoDB, Weka) *Course Project, Rutgers*, 2013
- Queried tens of millions of tweets to analyze the fractions of physically active people that live in different cities
- Designed a natural language analyzer (on *k*-NN) to identify physically active persons from those who just tweeting about exercise
- Visualized the geographic distribution of physical active users with heatmap, histogram, and leaderboard

Self-Educated Projects (projects briefly described with skills used)
› Discovering Urban Zonal Travel Demand Correlation using Points-of-Interest (POI) check-ins (Spatial-temporal data modeling)

SELECTED PUBLICATIONS (Full List)

1. Process Mining the Trauma Resuscitation Patient Cohorts 2018
Sen Yang, Fei Tao, Jingyuan Li, Dawei Wang, Shuhong Chen, Omar Z. Ahmed, Ivan Marsic and Randall S. Burd
Accepted by IEEE International Conference on Healthcare Informatics (ICHI 2017)

2. Automated Mining of Approximate Periodicity on Numeric Data: A Statistical Approach 2018
Ran He, **Sen Yang**, Jingyuan Yang, and Jin Cao
2018 ACM International Conference on Compute and Data Analysis (ICCDA2017)

3. A Data-driven Process Recommender Framework 2017
Sen Yang, Xin Dong, Leilei Sun, Yichen Zhou, Richard A. Farneth, Hui Xiong, Randall S. Burd and Ivan Marsic
2017 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2017)

4. Medical Workflow Modeling Using Alignment-Guided State-Splitting HMM 2017
Sen Yang, Moliang Zhou, Shuhong Chen, Xin Dong, Omar Z. Ahmed, Ivan Marsic, and Randall S. Burd
IEEE International Conference on Healthcare Informatics (ICHI 2017)

5. Process-oriented Iterative Multiple Alignment for Medical Process Mining 2017
Shuhong Chen, **Sen Yang**, Moliang Zhou, Randall S. Burd, and Ivan Marsic
IEEE International Conference on Data Mining Workshop (ICDM Workshop 2017)

6. VIT-PLA: Visual Interactive Tool for Process Log Analysis 2016
Sen Yang, Xin Dong, Moliang Zhou, Shuhong Chen, Ivan Marsic, and Randall S. Burd
KDD 2016 Workshop on Interactive Data Exploration and Analytics (KDD Workshop 2016)

7. Duration-Aware Alignment of Process Traces 2016
Sen Yang, Moliang Zhou, Rachel Webman, JaeWon Yang, Aleksandra Sarcevic, Ivan Marsic, and Randall S. Burd
Industrial Conference on Data Mining. Springer International Publishing, 2016

8. A Context-Aware Deep Learning Framework for Next Medical Treatment Activity Recommendation Under Review
Sen Yang, Weiqing Ni, Xin Dong, Leilei Sun, Shuhong Chen, Richard A. Farneth, Burd S. Randall, Ivan Marsic
Submitted to European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (PKDD 2018)

› My research is advised by Prof. Ivan Marsic and Prof. Hui Xiong (co-advisor).
› During my Ph.D. study, I have mentored and supervised 28 graduate and undergraduate students for their research credit (full list). Most of them are now working as software developers or data analysts in famous IT or finance companies.