

Updated in Sept. 2017

Fan Yang

• Cell Phone: (086) 131-1615-8612

• E-mail: fanyang0801@yahoo.com

• Hometown: Tianjin

CAREER PROFILE

- MS in Computer Science & MS in Statistics, plus 2 years experience as a statistician
- Solid programming skills in Java, Python, SAS, and R
- Knowledge and hands-on experience in machine learning algorithms and big data analytics tools
- Knowledge and experiences in statistical modelling and hypothesis testing
- Knowledge and project experience in web development
- Basic knowledge in finance and economics (passed CFA Level I)
- Linkedin: <https://www.linkedin.com/in/fan-yang-bbb919b1>

EXPERIENCE

Research Assistant

09/2016 - 09/2017

Computational Biomedicine Imaging and Modeling Center, Rutgers University

New Brunswick, NJ

- My research involves modeling of medical signals in Prof. Michmizos' group.

Statistician

05/2013 - 05/2015

Target Health Inc.

New York City, NY

- Reviewed clinical protocols & assisted in developing statistical analysis plan.
- Employed SAS procedures and developed SAS macros to conduct ad hoc statistical analysis per Clients' requests and FDA regulations.
- Experienced in data cleaning (outlier detection and multiple imputations)
- Got involved in frequent communications with clinical sites, central labs and large pharmaceutical companies (our clients).

Research Assistant

12/2007 - 06/2009

Institute of Orthopedics, Tianjin Hospital

Tianjin, China

- Publication: The Effect of Methotrexate on Synoviocytes Apoptosis in Pigmented Villonodular Synovitis. Fan XJ, Yang F, Xing GS, Zhao WJ, Lu Y, Bai RX. Tianjin Medical Journal, 2008 Issue 12.

EDUCATION

M.Sc. in Computer Science

09/2015 - 05/2017

Rutgers University – New Brunswick GPA 3.7/4.0

New Brunswick, NJ

- Thesis: Localization of Subthalamic Nucleus with K-means Clustering

M.Sc. in Statistics

09/2011 - 05/2013

Rutgers University – New Brunswick GPA 3.8/4.0

New Brunswick, NJ

- Waived master's qualifying exam

Ph.D. program in Biomedical Sciences, Drop Out

09/2009 - 07/2011

Oregon Health & Science University, School of Medicine

Portland, OR

- Activities: Vice President of Chinese Students & Scholars Association

B.Eng. in Biotechnology & Bioinformatics

09/2004 - 07/2008

Tianjin Medical University, School of Biomedical Engineering

Tianjin, China

CERTIFICATES

06/2015	Passed CFA Level I
12/2014	SAS® Certified Statistical Business Analyst Using SAS 9: Regression and Modeling (Certificate)
07/2011	SAS® Certified Advanced Programmer for SAS 9 (Certificate)
06/2011	SAS® Certified Base Programmer for SAS 9 (Certificate)
04/2006	Certificate for National Computer Rank Examination - Grade 3 (C language)

SKILLS & HOBBIES

Skills	<ul style="list-style-type: none">• Programming Languages: Java, Python, Shell Scripting• Math & Stat Tools: SAS, R, Matlab• Machine Learning Tools: NumPy, SciPy, Pandas, Scikit-Learn, Tensorflow, Mahout• Big Data Tools: Hadoop, MapReduce, Spark, Amazon Web Service (AWS)• Web Development Tools: HTML5/CSS3/Javascript, jQuery/AJAX, Bootstrap, PHP• Java-based Web Development: Java EE, Spring MVC, Hibernate, Tomcat 8.x• Database: MySQL, SQLite• Operating Systems: Windows, OS, Linux
--------	---

Updated in Sept. 2017

Languages Mandarin Chinese (Native) and English (Professional)

- Chinese English Test Band 6: 605/710
- Internet Based TOEFL: 106/120
- GRE General: V 640/800 (91%) + Q 800/800 (94%) + AW 4.5/6.0 (58%)

SELECTED PROJECTS

Master Thesis

- **Automated Localization of Subthalamic Nucleus using Modified K-means Clustering**

Sept. 2016 – Present

Advisor: Prof. Konstantinos Michmizos

Tools: Matlab, Python

The brain data was collected by microelectrode probes during deep brain stimulation surgeries. We filtered the 53 brain signals, i.e. the microelectrode recordings (MERs), from 24 patients with Parkinson's disease using a FIR filter with a band-pass between 0.5 kHz and 8 kHz. 18 computational features, either from the literature or from de novo design, were extracted from the filtered signals. After normalizing these features, we fed them into the K-means clustering algorithm and get the four segments corresponding to four regions of the brain on the trajectory of the probe. We further integrated the depth vector with the original feature matrix to lower the noise in the clustering results. Combining with the information in the feature activity maps, the K-means clustering can automatically identify the subthalamic nucleus boundaries on the MERs. Principal component analysis was also applied and reduced the complexity of the model.

Machine Learning

- **Scalable Logistic Regression Classifier Optimized by Stochastic Gradient Descent**

Oct. 2015 - Dec. 2015

Teammates: Bharath Joginapally, Sanjay Bharadhwaj Vijayaraghava

Tools: Java, Apache Mahout

We built logistic regression classifiers using Java and Apache Mahout, which were tested on data sets with increasing number of observations.

- **Network Intrusion Detection Using Machine Learning Algorithms**

Mar. 2016 - May. 2016

Teammates: Yang Zheng, Xin Cheng

Tools: Python, NumPy, Pandas, Scikit-Learn, Spark

We built an effective network intrusion detection system using different supervised learning algorithms, such as Naïve Bayes, decision tree, logistic regression, support vector machine, k-means clustering, and artificial neural networks.

(1) Stage 1 focused on binary classification of normal connections and cyber-attacks on 10 percent KDD cup 1999 data (~ 400 thousand training data). Packages like pandas and NumPy were used for data preprocessing. Scikit-Learn packages were employed for parameter optimization, kernel selection, and comparisons of different methods.

(2) Stage 2 focused on multiple classifications where (unsupervised method) k-means clustering and (supervised method) decision tree were implemented to further distinguish different types of cyber-attacks.

(3) The final stage involved k-means clustering using PySpark on the whole data (~ 5 Million observations).

Big Data Analytics

- **N-Gram based Google Auto-complete System**

Sept. 2016 - Oct. 2016

Tools: Java, Hadoop, MapReduce, jQuery, PHP, MySQL

The algorithm for this project involves two consecutive MapReduce processing. An N-Gram library was built from wiki database. On top of the N-Gram library, a language model was established based on the probability of phrase appearance. PHP and MySQL was employed for data query. jQuery and AJAX was used for front end presentation. Real-time auto-complete function was accomplished.

- **Ranking Websites using PageRank Algorithm**

Oct. 2016 - Nov. 2016

Tools: Java, Hadoop, MapReduce

Adjacency matrix was used to establish the relationships between different websites. The PageRank for different websites was computed. The convergence of PageRank algorithm led to ranking of websites.

- **Movie Recommender System**

Nov. 2016 - Dec. 2016

Tools: Python, Hadoop, Amazon EMR

A movie recommender system is built based on similarity between each user's past movie rating records. This system is capable of giving the most similar movies as top recommendations to each user. Two versions of this recommender system were developed using python: one took advantage of MRJob to work on Amazon Elastic MapReduce, the other was built for Spark. Both were run on a cluster of 20 nodes on the cloud. Results were analyzed and compared. The total number of observations in this project reached 20 million.

Web Development

- **A Spring-Hibernate CRUD Web Application of Student Registration System**

Dec. 2016 - Jan. 2017

Tools: Spring MVC Framework, Hibernate Framework, HQL, Java EE, JSP, HTML, CSS, XML, Tomcat 8.5

After initial setup of the development environment, Eclipse Java EE was connected with MySQL database using JDBC. A hibernate student entity class was then created and mapped to a database table in MySQL (Object-to-Relational Mapping). Hibernate session factory was utilized to create, read, update, and delete student objects upon user's request. Data access objective was later injected into controller. Modifications on the database were displayed to the user via a JSP page.