

Shijia Gu

(86) 186-1093-2730 ✦ kevin.sjgu@outlook.com

Floor 3, Building C, Area 16, CNPC Sci-tech Exchange Center, Changping, Beijing, China, 102200

EDUCATION

Certificates of Online Courses

DeepLearning.ai, Coursera, Big Data University etc.

Sequence Models (Mar. – Apr. 2018), Neural Networks and Deep Learning (Nov. – Dec. 2017),
Machine Learning (Oct. – Dec. 2013), Functional Programming Principles in Scala (Sep. – Nov. 2014), etc.

Master of Science in Engineering

Sep. 2009 – Mar. 2012

Beijing University of Posts and Telecommunications (BUPT), Beijing, China

Cumulative GPA: 3.32/4.0, Major Rank: Top 10%

Bachelor of Science in Electronic Information Science and Technology

Sep. 2005 – Jun. 2009

Minor in Computer Science and Technology

Nanjing University of Posts and Telecommunications (NJUPT), Nanjing, Jiangsu, China

Cumulative GPA: 3.72/4.0, Major Rank: Top 5%

EMPLOYMENT

Big Data R & D Engineer, Project Manager

Apr. 2015 – Present

Richfit Research Institute,

Beijing, China

China National Petroleum Corporation (CNPC) Beijing Richfit Information Technology Co., Ltd.

One of the first 4 founding staff, developing Big Data platform and machine learning algorithms.

Business Intelligence (BI) Development Engineer

Apr. 2012 – Mar. 2015

ERP Consulting Section,

Beijing, China

China National Petroleum Corporation (CNPC) Beijing Richfit Information Technology Co., Ltd.

Developed data models, transitions and reports on SAP and Oracle BI platforms.

PUBLICATIONS

Peer-reviewed and Conference

- **Shijia Gu**, Zhongxia Liang, Kang Di, and Boliang Zhao. “An Attention-based Tag Combination Approach for Multi-task Chinese Sequence Labelling Joint Learning”. (In preparation)
- **Shijia Gu**, Yueyang Song, Xin Zhao, and Weihai Li. “Fuzzing Test Data Generation based on Message Matrix Perturbation with Keyword Reference”. In 2011 IEEE Military Communications Conference (MILCOM2011), pages 1115 – 1120, Baltimore, USA, Nov. 2011.
- **Shijia Gu**, Weihai Li, and Xin Zhao. “Brute Force Vulnerability Testing Technology based on Data Mutation”. In 2011 IEEE Vehicular Technology Conference (VTC2011-Fall), pages 1 – 6, San Francisco, USA, Sep. 2011.
- **Shijia Gu**, Weihai Li, and Zhao Long. “Improved Vulnerability Testing Mode based on Syntax Analysis and Regular Expressions Test Cases Generation”. In Proceedings of IC-BNMT2010, pages 90 – 94, Beijing, China, Oct. 2010.

Patent

- **Shijia Gu**, Jun Zhu, Tao Wang, Yuanhang Zong, Zhendong Xu, and Chenxi Jia. “A Distributed ETL and Data Asset Management System and Method”. (Submitted)
- Weihai Li, **Shijia Gu**, Gang Liu, Yuqin Lv, and Xin Huang. “A Network Protocol Vulnerability Testing System and Method”. 2010. P.R. China, 201010267661.2.

PRESENTATIONS

Oral

“Research and Application of Big Data Technology in Petroleum and Natural Gas Industry” Jun. 2017
Seminar of Informatization Technology in Operation and Maintenance Projects Guiyang, Guizhou, China

“Fuzzing Test Data Generation based on Message Matrix Perturbation with
Keyword Reference”

Nov. 2011

2011 IEEE Military Communications Conference

Baltimore, MD

“Improved vulnerability testing mode based on syntax analysis and regular expressions
test cases generation”

Oct. 2010

2010 IEEE International Conference on Broadband Network and Multimedia Technology

Beijing, China

Poster

“Brute Force Vulnerability Testing Technology based on Data Mutation”

Sep. 2011

2011 IEEE Vehicular Technology Conference

San Francisco, CA

RESEARCH EXPERIENCES

Text Analysis Schemes based on Deep Learning (Richfit Youth Innovation Fund) Nov.2017 – Present
Research Institute, CNPC Beijing Richfit Information Technology Co., Ltd.

- Using deep neural network to explore better Chinese Word Segmentation and Named Entity Recognition.
- Searching similar phrase groups by semantic similarity pair-wise comparison with LSH algorithm.
- Extracting key information from raw contract text and classifying long encrypted documents.
- Standardizing short named entities with the same meanings.

Richfit Big Data Technology Platform (Richfit Scientific Research Project) Apr.2015 – Dec. 2016
Research Institute, CNPC Beijing Richfit Information Technology Co., Ltd.

- Constructed company's first 20-machine Spark and Hadoop cluster, including configuring and tuning.
- Read Apache Spark (version 1.0.0) source codes and analyzed its running mechanism.
- Developed applications and algorithms on Spark, used Spark SQL and MLlib in mass data modeling and analyzing.
- Reconstructed Spark and Hadoop RPM binary packages to fit in Apache Ambari management platform.
- Investigated Security Strategy Solution for Spark and HDFS.

Network Protocol Vulnerability Discovery (Chinese Universities Scientific Fund) Feb. 2010 – Nov. 2011
Intelligent Information Technology Laboratory, BUPT

- Assisted in proposing the vulnerability discovery methodology and designing the robustness testing framework.
- Wrote the detailed testing strategies and generated test cases using regular expression for HTTP, SIP and FTP.
- Introduced the XML as the protocol description and built up the testing platform including injector and receiver.
- Proposed a novel fuzzing data generation method based on matrix perturbation and a new protocol keyword extraction method with message semantic splitting.

Telecom User Group Information Mining System based on Cloud Computing Sep. 2010 – Dec. 2010
Research Institution of China Mobile

- Forecasted and classified customer churning with Logistic regression model.
- Assisted in clustering user group for the customer churning alerting system and analyzed the life trace of mobile users.

Signal Character Extraction, Identification and Electronic Counter Measures Technology Research in Network Cyberspace (National High-tech R&D Pre-program) Jun. 2010 – Oct. 2011
School of Electronic Engineering, BUPT

- Independently composed the application materials to win the bid and proposed the detailed research plan.
- Collected reference materials in Chinese and English and participated in the classification of blind separation algorithm.
- Assisted in constructing the simulation platform and evaluating selected separation algorithm under certain conditions.
- Listed outlines, guided and reviewed the two final research reports.

Speaker Recognition Technology Research on Deliberately Disguised Speeches Feb. 2009 – Jun. 2009
Speech and Intelligent Signal Processing Laboratory, NJUPT

- Recorded 8 kinds of disguised voice and normal speech of 10 people in acoustic chamber with special designed corpus.
- Extracted features from raw speeches and tried to evaluate different pattern recognition models based on the signals.
- Compared the parameters when trained Gaussian mixture models with normal speech with different initial value.
- Estimated the speaker recognition performance on disguised voices in 10 people group.

INDUSTRIAL EXPERIENCES

CNPC Joint Supervision Big Data Project Nov. 2016 – Present

- Searching illegal oil card arbitrage and regular off-line sale based on gas station transaction records.
- Analyzing underselling below purchase costs or daily average prices in convenient stores.
- Detecting illegal tendering and bidding based on vendors and purchasing information.
- Identifying affiliated and similar contracts on the basis of contracts text.
- Proposed a novel CWS and NER combination method to improve contract texts segmentation.

Distributed ETL on Apache Spark and Data Asset Management Tool Jan. 2018 – Oct. 2018

- Led a six-member team, designed software architectures and defined functional details.
- Developed a parser to generate back-end jobs according to configuration file, gave advises on problems in programming.
- Added data lineage retrieval and visualization module based on Neo4j, a graph database.
- Developed data asset management tool for multi-level management and maintenance.
- Composed documents to apply software copyright, including user handbooks and source codes samples.

CNPC Security Operation Center

Apr. 2016 – Oct. 2016

- Preprocessed seven days' tera-byte netflow packages, removed useless or invalid data.
- Designed 7 abnormality features to judge botnet IPs by clustering, frequent pattern mining and statistical methods.
- Used wavelet analysis to detect malicious network scanning hosts in real time.

Well Site Real-time Online Data Processing Platform

Dec. 2015 – Feb. 2016

- Constructed and trained a model to predict Circulation Loss and Sticking accidents with "Price Rate" Algorithm.
- Assisted in developing a real-time data analysis software to monitor and alert on potential drilling accidents.

CNPC Equipment Company Decision Support System

Oct. 2014 – Mar. 2015

- Built data objects and cubes to compute production index values and showed them in dashboard reports.

CNPC Transportation Company Business Warehouse and Report System

Aug. 2012 – Sep. 2014

- Developed new logic models, fixed bugs, uploaded data monthly from ERP to BW system.

CNPC Dagang Oilfield Report Acceleration Pilot Project

Apr. 2012 – Jul. 2012

- Migrated data in traditional warehouse system to in-memory database, redesigned and optimized data models.

TEACHING EXPERIENCES

Tutor of New Staff Training

Jul. 2016 & 2017

Research Institute, CNPC Beijing Richfit Information Technology Co., Ltd.

- Lectured on Apache Spark and Functional Programming.
- Designed and supervised on cluster deployment and application development.

Tutor of Customer Training

Mar. 2013

Beijing Control Center, CNPC Transportation Company Limited

- Lectured on SAP BW report system and illustrated business logics and technical processes.
- Instructed on usages of monthly and annual statistical reports.

Assistant Tutor of Bachelor Thesis

Feb. 2011 – Jun. 2011

Teaching and Researching Center of Electric Circuit and System, BUPT

- Instructed and supervised weekly on two students' graduation projects and theses.

Director of Laboratory Academic Seminar

Sep. 2010 – Mar. 2012

Intelligent Information Technology Laboratory, BUPT

- Held bi-weekly discussion on selected papers from top conferences and journals to share innovative ideas.

AWARDS AND HONORS

A-grade Employee of the Year (Top 5%)

Dec. 2016 & 2017

Champion of IBM Big Data Hackathon

Aug. 2015

Excellent Graduate Student

Mar. 2012

Travel Grants by Both BUPT and MILCOM

Nov. 2011

Second-class National Scholarship and Tuition Waiver (Top 5%)

Sep. 2010 & 2011

Third-class National Scholarship (Top 10%)

Sep. 2009

Excellent Graduation Thesis

Jul. 2009

Third-class NJUPT Scholarship (Top 10%)

Nov. 2007

Second-class NJUPT Scholarship (Top 5%)

Nov. 2006

COMPUTER SKILLS

Computer Languages

Scala, Python, Java, C/C++, MATLAB

Operating Systems

Linux, Windows

SoftwareKeras, Tensorflow, Spark, Hadoop, L^AT_EX, MS Office**EXTRACURRICULAR ACTIVITIES**

Member of Company Badminton Team

Apr. 2012 – Present

Member of School Badminton Team

Dec. 2009 – Mar. 2012

Member of School Basketball Team

Sep. 2005 – Aug. 2008