

EDUCATION

- Renmin University of China**, *Master of Laws in Sociology*; GPA 3.65, Beijing, China Sep 2015 – Jul 2017
- Relevant Coursework: Data Analysis and Using STATA (A); Advanced Social Statistics (A); Application of SAS Software (A)
- China Agricultural University**, *Bachelor of Laws in Sociology*; GPA 3.77, Beijing, China Sep 2011 – Jul 2015
- Relevant Coursework: Advanced Math (90); EXCEL and Data Analysis (94); Social Statistics (91); Social Statistics Analysis Method (100); Quantitative Research (100)

PROFESSIONAL EXPERIENCE

- Tianjin Aden Tech Co., Ltd.** – *Senior Data Analyst, Department Data& Algorithm*; Beijing, China Feb 2019 – Mar 2020
A hospital management and consulting company that provide services for a large number of hospitals (3000+) and institutions in China
- Supported consulting efforts to attract major clients, like **PING AN INSURANCE COMPANY OF CHINA, National Health Commission**
 - Built up & maintained a data analysis platform based on **Hadoop & Spark** that can support fast & real-time aggregation operation of 100,000,000 of data points
 - Processed & analyzed 3000+ hospitals' data using **Python**, and conduct data **ETL**(extract, transform, load) and index computing
 - Assisted consulting team to design hospital performance management system prototype and coordinated with developing team for all project implementation
 - Worked with insurance team to develop payment model based on optimization method (**Pandas, NumPy, Scikit-learn**)
 - Processed text data and increase the medical words matching accuracy by 20% using **NLP method**

- Tianjin Aden Tech Co., Ltd.** – *Data analyst, Department Data& Algorithm*; Beijing, China Jul 2016 – Feb 2019
- Optimized Diagnosis Related Groups (DRGs, a medical insurance payment model) results using **Random Forest algorithm**. Achieved 25% better accuracy vs traditional version (**Hive, Pyspark**)
 - Devised hospital inpatient records quality inspection system that improve the rule writing efficiency by 50%
 - Increased inpatient records quality inspection system operation speed by two orders of magnitude (100x Faster) using **Scala**
 - Visualized collected 31 provinces data using **Tableau** and **Echarts** (a visualization web tool)

PROJECT EXPERIENCE

- YALE University - Retiree Health Status Research** Jul 2019 – Jun 2019
- Built and optimized feature engineering pipeline with Data Acquisition (**Mysql**), data processing (**Pandas**), feature encoding (**Scikit-learn**)
- Social Integration of Migrant Laborers Research** Jul 2016 – Jun 2017
- Collected and Analyzed 5 migrant laborers data sets using **Python/Stata**
 - Applied a **Time Series Model & Logistic Regression Model** to analyze factors affecting social integration

HONORS AND ACADEMIC MEMBERSHIPS

- Fellow, International Society for Quality-of- Life Studies (ISOQOL) Aug 2020
- Outstanding Graduate of Beijing Jun 2015
- National Scholarship (awarded for the top 5% students) Jun 2014

SKILLS AND INTERESTS

Computer Skills: Programming in Python, R, Scala, SAS, Stata, SPSS; Tableau; SQL; Hadoop, Hive; Spark; Git
Languages: English (Fluent), IELTS: 6.5, Mandarin Chinese (Native)
Interests: Chinese martial arts (3 years), Kendo (1 year), Swimming

PUBLICATIONS

- **G. Sun**, L. Cheng, Q. Su, *Random-forest-algorithm-based method and system for identifying disease diagnosis complication*, China Patent No. CN108376564A
- **G. Sun**, L. Cheng, W. Zhu, *Medical data statistical query system and method*, China Patent No. CN110941650A
- **G. Sun**, L. Cheng, Q. Su, *Method and system for carrying out automatic disease code conversion on basis of semantic approximate matching algorithm*, China Patent No. CN108446260A
- **G. Sun**, L. Cheng, W. Zhu, *Medical corpus labeling method*, China Patent No. CN110826304A

孙广阳

<https://user4gith.github.io/>

yang4comm@hotmail.com

(86) 130-5158-5963

教育背景

中国人民大学, 社会学硕士 (保送研究生); GPA 3.65

Sep 2015 – Jul 2017

- 相关课程: 数据分析及STATA软件应用 (A); 高级社会统计方法 (A); SAS统计软件应用 (A)

中国农业大学, 社会学学士; GPA 3.77

Sep 2011 – Jul 2015

- 相关课程: 高等数学 (90); EXCEL和数据分析 (94); 社会统计学(91); 社会统计分析方法 (100); 定量研究 (100)

工作经历

艾登科技有限公司 – 高级数据分析师, 数据与算法部

Feb 2019 – Mar 2020

医院管理和咨询公司, 在中国为超过3000家的医院和相关机构提供服务, 客户包括平安保险和卫生健康委员会

- 为咨询团队提供核心数据分析支持, 服务客户包括中国平安保险, 国家卫生健康委员会
- 搭建和维护基于大数据Hadoop和Spark技术的数据仓库, 实现亿级数据的快速运算
- 处理和分析3000+医院的数据, 包括数据ETL(extract, transform, load), 指标设计, 协同咨询团队合作出具医院绩效报告
- 协助咨询团队设计医院绩效管理和病案质量管理系统的分析模块, 并协调开发团队梳理需求, 协助项目实施
- 同保险团队合作开发基于最优化算法的保险支付模型(Python, Pandas, NumPy, Scikit-learn)
- 协助医学团队进行病种探索性分析, 基于图数据库Neo4J
- 处理技术处理和分析文本数据, 利用自然语言处理技术提升医学文本转换疾病代码准确率20%

艾登科技有限公司 – 数据分析师, 数据与算法部

Jul 2016 – Feb 2019

- 优化疾病诊断相关分组 (DRGs, 医疗保险支付模型) 结果, 利用随机森林算法对原有分组结果进行优化, 将分组结果准确率提升25% (Hive, Pyspark)
- 开发病历质量检查系统, 实现对病历质量的快速判断, 利用Scala将运行速度提升两个数据级 (100毫秒提升至1毫秒)
- 开发医学规则管理系统, 设计医学规则书写模型, 提升规则书写效率50%
- 可视化分析收集的31个省级行政单位数据, 利用Tableau和Echarts可视化方案

项目经历

耶鲁大学 – 退休人员健康状况研究

Jul 2019 – Jun 2019

- 建立和优化特征工程 pipeline 包括数据收集 (Mysql), 数据预处理(Pandas, NumPy), 特征编码 (Scikit-learn)

打工人员社会融合研究

Jul 2016 – Jun 2017

- 收集和处理5个流动人口国家级数据集 Python/Stata
- 利用时间序列和逻辑回归模型分析影响社会融合的因素

荣誉

- 协会成员, International Society for Quality-of- Life Studies (ISOQOL) Aug 2020
- 北京市优秀毕业生 Jun 2015
- 国家奖学金 (前5%的学生) Jun 2014
- 一等奖学金 (前10%的学生) Jun 2013

技能爱好

电脑技能: Python, R, Scala, SAS, Stata, SPSS; Tableau; SQL; Hadoop, Hive; Spark; Git

语言: 英语 (流利), IELTS: 6.5, GRE: Verbal: 152, Quantitative: 164, Analytical writing: 3.5

爱好: 散打 (3 年), 游泳

公益服务: 流动儿童成长计划志愿者 (Sep 2012 – Present)

专利

- G. Sun, L. Cheng, Q. Su, Random-forest-algorithm-based method and system for identifying disease diagnosis complication, China Patent No. CN108376564A
- G. Sun, L. Cheng, W. Zhu, Medical data statistical query system and method, China Patent No. CN110941650A
- G. Sun, L. Cheng, Q. Su, Method and system for carrying out automatic disease code conversion on basis of semantic approximate matching algorithm, China Patent No. CN108446260A
- G. Sun, L. Cheng, W. Zhu, Medical corpus labeling method, China Patent No. CN110826304A