

黄志翔

应聘岗位：大数据研发工程师
数据挖掘工程师



个人信息

D.O.B. 1995.9.20
18482202023
515791278@qq.com



自我评价

- ◆ **教育背景**：本科一本专业在读，通过国外公开课自学大数据、数据挖掘，自然语言处理等学科，专业知识完备。
- ◆ **团队协作**：优秀的团队合作者，有团队开发经验，实验室相处融洽，在校期间积极参与辩论赛，社团等活动。
- ◆ **学习能力**：能够快速学习新知识，并应用于实践，长时间通过 YouTube，Coursera 等公开课自学多个专业。
- ◆ **性格爱好**：阳光乐观、目标明确、抗压力强、热爱计算机，大数据；爱好运动、健身，吉他。



教育及工作背景

2014.09-2018.07	西华大学	物流管理	本科（在读）
2017.01-2017.08	大数据实验室（本校）	大数据	科研
2016.06-2016.09	勤智数码	软件开发	实习



专业技能

英语水平：CET-4，熟练阅读英文 API，原版书，学习多门国外公开课，满足日常口语交流。

基础技能：1.熟练使用 Java，Python 开发，熟练使用 SQL，简单掌握 Scala，Shell 脚本,了解 Lisp，C 语言。
2.熟悉计算机原理，JVM，汇编。
3.了解多种开发设计模式，对数据和需求敏感。
4.熟悉 Linux 操作系统，远端服务器，SPSS 数据分析软件。

数据挖掘：

- 1.熟悉中文分词(多个开源库及原理)，结构化数据，数据清洗，数据降维，特征工程，模型优化与调参。
- 2.熟悉深度神经网络，对（RNN，LSTM 有研究），熟悉 TensorFlow 框架，理解 Word2vec。
- 3.实现基础数据挖掘模型，如朴素贝叶斯，KNN，SVM，K-means 聚类，决策树，协同过滤等。

大数据研发：

- 1.独立完成集群环境的部署，了解 ZooKeeper 分布式协调服务的使用和原理。
- 2.了解 HDFS、MapReduce、yarn 等 Hadoop 生态体系技术的使用和原理，编写 MapReduce 程序。
- 3.了解 Hive,Hbase 的存储原理，能够使用 HQL 进行数据分析，实现定时执行的离线系统。
- 4.了解 Flume+Kafka+Storm 流式计算内部原理，实现简单需求。
- 5.掌握 Redis 实时 NoSQL 数据库的使用
- 6.能够使用 Sqoop 进行关系型数据库与 HDFS/Hive/MySQL 之间进行数据导入导出。
- 7.掌握 Scala 语言基础，可以使用 Scala 进行简单编程。
- 8.了解 Spark 核心源码实现，对 RDD 编程，Spark-Sql 有一定的了解与使用经验。

Python :

1. Python 基础扎实，多采用过程式，函数式编程。
2. 实现和研究多个机器学习算法，使用 BeautifulSoup，布隆过滤器等编写爬虫。
3. 使用 NumPy，Pandas 操作数据并计算，使用 Matplotlib 绘制准确率，词云等。
4. 熟悉 Scikit-learn 算法库，熟悉 Google 开源框架 TensorFlow 构建 LSTM 等深度学习模型。

Java EE :

1. Java 基础扎实，熟悉多线程，Socket，JDBC，反射，IO 流等技术，有良好的编码风格。
2. 熟练运用 Spring，SpringMvc，Hibernate，Mybatis，Shiro 框架。
3. 了解 Spring 中 NIO，RPC 机制，其他框架内部设计模式。
4. 熟悉 HTML，CSS，JSP，Servlet，JSON，XML，等 J2EE 相关技能。
5. 了解 JavaScript，AJAX，jQuery 等客户端相关技术。
6. 熟悉 Tomcat 服务器，JUnit 单元测试工具，简单的 debug 调试。

其他 :

1. 熟悉 Linux 服务器环境搭建及项目部署、及常用 Shell 命令，熟练使用 Xshell。
2. 熟悉 SVN，Eclipse，Maven，PyCharm 等的使用。
3. GitHub 地址：<https://github.com/515791278>

实习经历

2017.01-2017.08

大数据实验室

科学研究

项目一：疾病智能化识别系统（自然语言处理）

使用语言：Python+MySQL

研究课题：基于 23 万份历史病例（成都某三甲医院），智能诊断（预测）疾病，提取疾病特征词，量化医师技术水平

开发过程（参与整个流程）：

一期（团队）：机器学习

1. 项目文档及架构，操作医院 HIS 系统，从 SQLite 数据库中拉取信息。
2. 数据前期清洗，统计疾病总数（SQL 及聚类），确定目标疾病。
3. 中文分词研究，分词词典，停用词表构建。
4. 特征工程，采用 TF-IDF, Word2vec 降维，结合人工提取特征（主要）。
5. 研究算法原理并运用，如朴素贝叶斯，K-means 聚类，TF-IDF 降维，KNN，SVM，Decision tree 等算法。
6. 使用 Scikit-learn 算法库，Pandas，Matplotlib，最后基于 KNN, Naive Bayes, Neural network 组合构建模型。
7. 调整参数，优化和提取特征，预测疾病的准确度从初始 46% 提升到 79%。自动识别部分疾病特征。
8. 学习 TensorFlow，Word2vec，搭建 RNN, LSTM 语言模型，帮助医生填写病例（自动文本提示）。

二期（自学与应用为主）：医疗数据仓库

1. Linux 下搭建 Hadoop 集群，使用 HDFS 存储数据，MapReduce 简单统计词频。
2. 离线部分：原始医疗数据，使用 Sqoop 导入 Hive（分区存储），Flume 实时拉取数据到 HIVE 仓库，编写 HQL，制作医院报表，将分析数据输出到 MYSQL，设置 crontab 定时执行 Shell 脚本。
3. 实时计算部分：使用 zookeeper 管理集群，Flume 实时拉取数据，通过 Kafka 传输到 Storm 流式计算，结果保存到 Redis。
4. 练习 Scala 语言以及编写 Spark 程序和 Spark-Sql。

2016.06-2016.09

勤智数码（150-500 人）

协助企业开发 实习

项目二：医学标本实验预约系统

软件架构：Spring+SpringMvc+Hibernate+Shiro+MySQL+JSP+ MySQL+jQuery

开发环境：STS + JDK1.7 + MySQL+ Maven+ SVN

项目描述：向医院师生开通在线预约功能，标本库工作人员根据申请情况和设备实际使用情况，进行预审，优化排程，提高设备使用的效率，并通过短信平台提高沟通效率。

责任描述：用户管理模块，用户登录，耗材管理，角色管理树形菜单。

开发详情：

1. 根据 UML 模型图建立自己模块的 Hibernate 图表关系。
2. 实现模糊查询，在公司搭建好的框架上进行开发。
3. 熟练使用 MVC 模式，根据业务需求编写代码。
4. 使用 SVN 管理项目进行团队协作开发。
5. 通过 AOP 使用 Log4j 进行日志管理，Shiro 权限控制。

实习心得：

1. 了解整个项目的工作流程，学会按需求开发，团队的沟通与协作，命名规范及项目计划分工。
2. 在项目经理的指导下，学习了如何通过抽象类，接口来减少冗余代码。
3. 深入学习了框架中的设计思想，如抽象工厂，单例，观察者，代理等设计模式，接触了 Nio, rpc 的框架。

获得奖励及职位

2017.07

《数字化用户》录用《医疗大数据的应用和数据仓库的设计》

2017.04

国家中级健身教练资格证

2016.03

获得 2015 年度院级奖学金

2015.09

担任创新创业俱乐部部长

2014.10

担任院级辩论队一辩