

大数据岗位人才需求分析与可视化

刘思思

(广东科技学院计算机学院, 广东 东莞 523083)

摘要: 为了深入了解大数据企业、岗位及人才需求的相关情况, 本文通过网络爬虫爬取“前程无忧”大数据岗位的相关招聘信息, 并对数据进行预处理和分析, 通过可视化技术从多个维度直观的展示大数据相关企业和岗位对人才的需求情况, 在一定程度上能够给大数据专业的毕业生提供一定的参考建议和就业指导。

关键词: 网络爬虫; 招聘数据; 数据分析; 数据可视化

据统计, 近年来, 人工智能、大数据等行业招聘需求爆发, 大数据行业 2019 年新发职位同比增长 16.66%, 2021 年较 2020 年同比增长 32.57%。然而, 目前大数据行业存在着人才缺口大、行业分布不均的特点, 数据显示, 预计到 2025 年, 大数据核心人才缺口将高达 230 万人。可见, 当今人才市场面临毕业生找不到理想工作、企业招不到理想员工的双重困境, 呈现出结构性失业严重的现状。

各类招聘网站的兴起为研究者分析招聘信息提供了渠道, 为了方便求职者在海量的数据信息中快速定位对自己有用的信息。本文爬取了大数据专业相关岗位的招聘信息, 对企业的招聘需求进行分析, 为大数据专业学生就业选择提供积极的指导意义。

一、系统设计

本系统基于 Python 实现, 主要分为数据爬取、数据存储和预处理、数据分析和数据可视化四个模块。整体框架如图 1 所示。

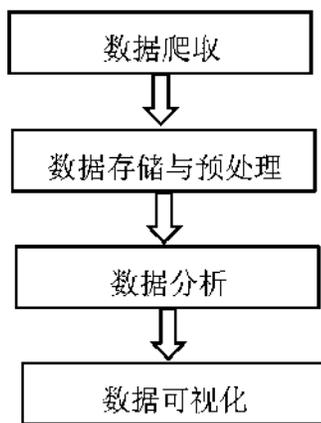


图 1 系统整体框架

(一) 数据爬取。以“数据”, “大数据”和“Python”三个关键词爬取“前程无忧”网站上的相关招聘信息, 采用 Python 语言的 Selenium 模块和 BeautifulSoup 模块来实现。首先利用 Selenium 库模拟登录前程无忧网站, 并输入关键词进行搜索翻页等操作, 获取 HTML 网页信息, 然后利用 BeautifulSoup 库精确解析 HTML 文档, 获得需要抓取的信息, 包括: 职位名称, 发布时间, 薪水, 公司地点, 工作经验要求, 学历要求, 招聘人数, 公司名称, 公司类型, 公司人数, 工作类型等。

(二) 数据存储与预处理。本文在爬取数据时, 使用了三个关键词, 不可避免的会存在一些重复的、缺失的, 错误的或不符

合规范的数据, 因此, 在进行数据分析之前, 需要对这些“脏数据”进行预处理, 本文使用 Pandas 模块进行数据的预处理工作。首先使用 pandas.read_excel() 将数据保存为 DataFrame 结构, 使用 pandas.info() 查看数据的整体情况, 将与分析无关的数据列删除。对缺失值的处理需要依情况而定, 比如对于“学历要求”和“工作经验要求”的缺失值可以替换为“不限”, 对于“薪水”的缺失值, 在缺失值不多的情况下可以选择删除。DataFrame.drop_duplicates() 可以快速按行检测并删除重复的记录。对于数据中表示不规范的数据进行转换或拆分, 最后将得到的 2291 条干净数据保存到 excel 或 MySQL 数据库中。

(三) 数据分析。数据分析是本文的关键与核心, 将在数据预处理过程中得到的数据按照具体的分析需求进行数据的分类聚合和筛选, 以备后续可视化使用。本文使用 Pandas 模块对数据进行分析和处理。主要包括了以下几个方面的分析: 各城市的岗位需求、各城市的平均薪资、经验要求及不同经验要求的平均薪资对比、学历要求及不同学历要求的平均薪资对比、热门职业的词云图。

(四) 数据可视化。数据可视化是使用图表的形式直观的展示数据分析的结果, 本文使用 pyecharts 开源可视化库绘制柱形图、饼图、词云图等图形展示大数据相关岗位的信息。

二、可视化结果展示与分析

(一) 全国各地大数据岗位需求

从分析和可视化结果可以看出, 大数据岗位的人才需求主要集中在北上广深等一线或新一线大城市, 全国大数据开发需求和供给主要也这些城市为主。因此大数据人才也倾向于流向这些地区。

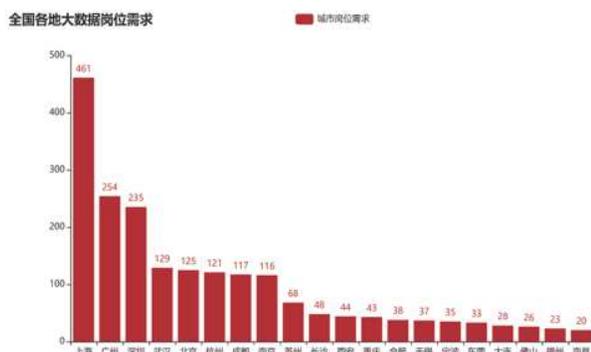


图 2 全国各地大数据岗位需求

(二) 学历要求和工作经验要求

当今社会, 竞争压力越来越大, 社会对人才的要求越来越高, 对学历也越来越重视, 本科学历是大数据相关职位的一个基本要求, 从各个职位上看需求量都是最大的, 高端人才要求硕士以上的学历也很普遍, 从分析结果也可看出, 本科学历就占了 67.61%。

同时大数据岗位对工作经验非常的看重，从图 4 结果可以看出，绝大部分岗位要求具有 3-4 年的工作经验，须需经验的所占的比例非常的小，这对应届毕业生来说压力比较大。数据分析师、数据挖掘师这些职位对行业知识和业务理解能力要求较高，需要一定的经验积累。

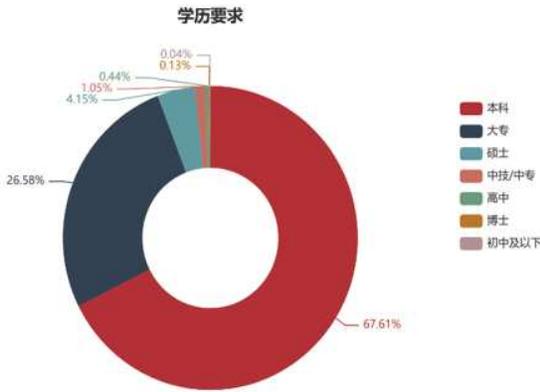


图 3 学历要求



图 4 工作经验要求

(三) 不同学历、不同经验的平均薪资对比

从上面的分析，大数据岗位对学历的要求及工作经验的要求都较高，说明大数据行业的进入的确是有一定门槛，但同时，大数据行业的薪资也是普遍较高的，平均月薪能够在 15K-20K 左右，且与学历和工作经验呈正相关的关系。由图 5 图 6 可以看出，本科学历的平均工资在 17K 左右，硕士学位为 20K，博士学位的高端人才的平均薪资达到了 32K。同时，平均月薪与工作经验也是呈线性增长，10 年以上工作经验的平均月薪高达 40K。

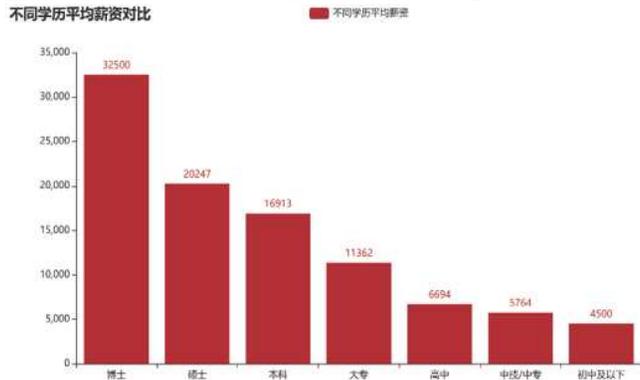


图 5 不同学历平均薪资对比

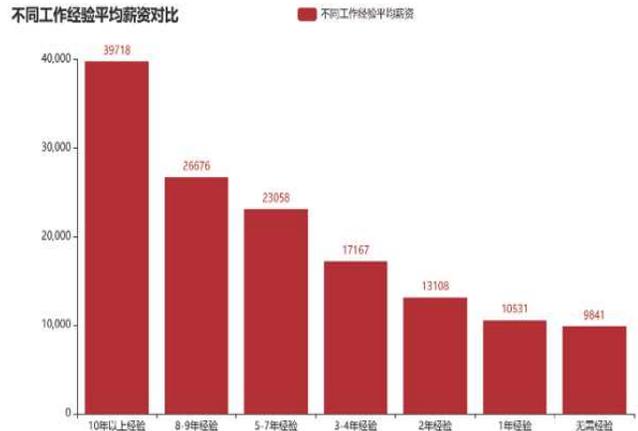


图 6 不同经验平均薪资对比

(四) 热门职位

随着人工智能和大数据技术的发展，大数据渐渐成为继计算机之后又一火爆专业，在国家经济发展和相关政策的推动下，大数据发展势头十分猛烈，经过几年的发展，已经形成了一定的产业链。大数据的就业方向主要有三，分别是研发类、数据分析类、应用开发类，各个方向对应的基础岗位，分别是大数据系统开发工程师、大数据应用开发工程师、大数据分析师。



图 7 大数据热门职位

四、结语

本文基于 Selenium 和 BeautifulSoup 模块爬取了“前程无忧”招聘网站上的“大数据”专业相关的招聘信息，并对数据进行分析和处理，以各种图表直观地展示了大数据的热门职位、城市的岗位需求、对学历和工作经验的要求以及不同学历和工作经验的平均薪资的对比，为大数据专业学生的求职选择提供一定的建议和便利，具有一定的社会价值。

参考文献：

[1] 扶明亮, 李刚. 基于招聘网站大数据分析的求职者需求与企业需求匹配研究 [J]. 统计与管理, 2021, 36 (08): 91-97.

作者简介：刘思思 (1986-)，女，广东科技学院，讲师，主要研究方向为数据分析与数据挖掘、机器学习。