

文章编号: 2095-2163(2020)04-0276-04

中图分类号: TP311.1; G353.12

文献标志码: A

本科层次大数据人才需求可视化分析

梁婷婷, 邱素贞, 陆珊

(南宁学院 信息工程学院, 南宁 530200)

摘要: 本文爬取2020年3月大型人才招聘网站的十几万条关于大数据人才真实招聘信息, 筛选出教育背景要求为本科学历(占比最大)的有效记录, 采用描述统计分析方法, 研究当前大数据人才的地域分布、薪资水平、经验要求、岗位要求等; 结合词频分析和统计技术, 挖掘出各企事业单位对大数据人才的素质能力、知识储备等要求。分析结果以可视化图形展示, 并提出四点建议, 对各本科高校制定大数据相关专业的人才培养方案具有实际意义。

关键词: 大数据; 人才需求; 数据分析; 可视化

Visual analysis of undergraduate level big data talent demand

LIANG Tingting, QIU Suzhen, LU Shan

(School of Information Engineering, Nanning University, Nanning 530200, China)

[Abstract] There is a strong demand for big data talents in the market. In March 2020, we crawled through the hundreds of thousands of real recruitment information about big data talents on the large-scale recruitment website, screened out the effective records with the education background of undergraduate degree (the largest proportion), and used the descriptive statistical analysis method to study the current regional distribution, salary level, experience requirements, job demand, etc. of big data talents. Combined with word frequency analysis and statistical technology, the requirements of enterprises and institutions on the quality, ability and knowledge reserve of big data talents are explored. The analysis results are displayed in visual graphics, and four suggestions are put forward, which is of practical significance for colleges and universities to formulate talent training programs for big data related majors.

[Key words] big data; talent demand; data analysis

0 引言

根据赛迪发布的《2019—2021年中国大数据市场预测与展望数据》显示, 预计未来三年中国大数据市场规模仍将保持30%以上的增长速度^[1]。到2021年, 中国大数据市场规模将达到4920.3亿元。希捷科技的一项最新研究表明, 到2025年全球数据量将会从2018年的33ZB上升至175ZB^[2]。到2020年, 工业机器人间分享的数据将远远超过人类所生成的数据量。爆炸式增长的数据量推动和丰富了大数据分析及应用需求。各地数字经济和新型智慧城市建设项目需要大量的数据人才。为了应对大数据人才的旺盛需求态势, 国家加快设立了一批相关专业。根据教育部公布的2017—2020年普通高等学校本科专业备案和审批结果显示, 获批新增“数据科学与大数据技术”专业的高校累计达到616所。此外, 国家还加快设立了交叉性的新学科, 培育既懂大数据技术, 又有专业知识(如智能建造、智能

医学、智能车辆、计算金融等)的复合型人才。

人才的培育顺应国家政策与当前及未来的市场需求。收集具有广泛影响力的人力资源服务供应商的人才招聘网站数据, 研究海量真实的名企职位招聘信息, 从当前的市场实际需求, 定位人才的素质能力、知识储备等要求, 对各高校制定人才培养方案具有实际意义。

1 研究数据及方法

本研究以“大数据”为关键词, 于2020年3月爬取前程无忧51Job人才招聘网站的52403条信息。其中包括12个字段: 岗位名称、薪资、教育背景、经验要求、招聘人数、公司名称、工作地区、公司性质、公司规模、所处行业、发布日期、职位详情等。去除重复、大量的房产销售等人才招聘广告的噪声数据, 以及一些字段数据为空的记录, 留取22531条有效数据。为了深入分析人才需求定位, 利用tableau数据可视化分析软件, 采用描述统计分析方

基金项目: 广西高等教育本科教学改革工程立项项目(2019JGB418)。

作者简介: 梁婷婷(1983-), 女, 硕士, 副教授, 主要研究方向: 信息检索、计算机应用技术; 邱素贞(1982-), 女, 硕士, 副教授, 主要研究方向: 计算机技术、信息化教育; 陆珊(1983-), 女, 学士, 高级工程师, 主要研究方向: 计算机技术。

通讯作者: 邱素贞 Email: 252718055@qq.com

收稿日期: 2020-02-03

法,研究当前大数据人才市场的实际需求,包括地域分布、薪资水平、经验要求、岗位需求等。结合词频分析和统计技术,经过对不同岗位的详情描述进行文本分析,挖掘出各企事业单位对大数据人才的素质能力、知识储备等要求。通过散点图、条形图、地图等可视化图形展示了主要分析结果。

2 研究结果及分析

本次爬取的数据中,有一万余条(占总数 20%)记录的招聘人数为“若干人”。为了更好的统计招聘人数,将招聘人数为“若干人”转换为 5 人。最终计算出总的招聘人数为 75 876 人。

企事业单位发布招聘信息时间分布在 2020 年 1—3 月份,3 月份发布的招聘人数占 74%,需求人数激增。分析原因,应是受鼠年春节和新冠病毒疫情影响,企业需求意愿不强。

数据显示,教育背景要求本科的占比最大,达到 63.4%,大专 20%和硕士 5.2%。可见本科学历的人

才最受欢迎,本科层次人才培养对整个大数据市场具有重要意义。为了更好的研究本科层次大数据人才的需求状态,本文筛选出教育背景明确要求为本科学历的记录进行研究,从多角度分析并给出相应结果。

2.1 岗位需求

各用人单位填写的招聘岗位名称不尽相同,部分有交叉重复,共有 6 579 种不同名称。经分析,结合大多数实际岗位的工作内容,将这些岗位名称进行分组整理。分组时,会有部分岗位名称涉及不同的组别(如:安全研发架构师),但是仅能处于一个组别。如图 1 所示,需求人数超过总体平均值的有:数据分析(含数据分析、统计分析、数据统计、财务分析、运营分析、数据专员等统一归为数据分析)、数据开发、软件开发方向、研发方向、数据工程师、大数据运维、运维方向、架构方向、售前&销售、算法方向、测试方向、解决方案、数据产品经理、DBA 数据库管理、数据挖掘方向、高级工程师、科学家...、项目管理、安全方向、产品经理、设计、AI 方向、应用、爬虫&数据处理方向、讲师和培训、智慧、通信、网络、物联网、数据平台、BI 方向、数据管理、数据建模&数据模型、可视化、财务方向等。

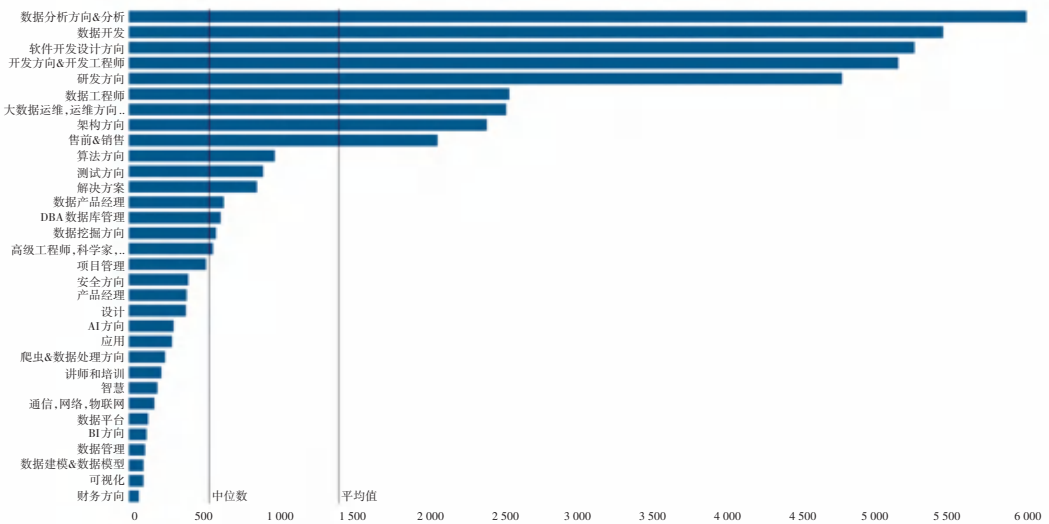


图 1 不同岗位需求人数条形图

Fig. 1 Bar chart of the number of people in different positions

2.2 薪资水平分布

各用人单位给出的薪资范围千差万别。有元/小时、元/天、千/月、万/月、万/年等单位,以及数据类似于“6.5~20 千/月”这样的格式,数据范围跨度不一。为了方便统计分析薪资水平,将所有的薪资单位换算为千/月,并将每个薪资数据范围分离出最低值和最高值,进而计算出每个薪资数据的平均值。经过统计,大数据人才的平均月薪为 17 千/月。并将薪资数据每 5 千为一个数据项绘制平均薪资水平分布直方图。如图 2 可见,大多岗位的平均薪资分布在 5~20 千/月。

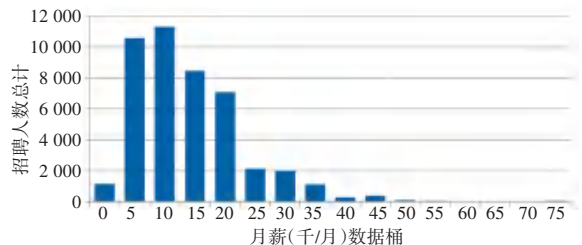


图 2 平均薪资水平分布直方图

Fig. 2 Distribution histogram of average salary level

对数据进一步钻取信息,绘制不同岗位招聘人数及工资散点图,如图 3 所示。可见,招聘人数比较多的岗位,如:数据分析、数据开发、软件开发等,平

均薪资在15千/月左右。而架构师、BI和智慧方向的岗位薪资比较高,达24千/月以上。

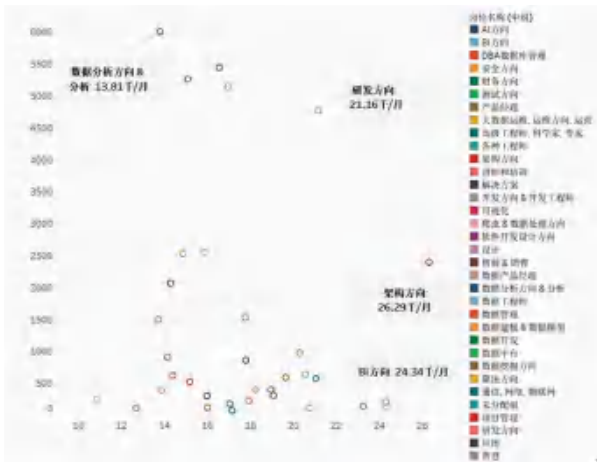


图3 不同岗位招聘人数及工资散点图

Fig. 3 recruitment number and salary scatter of different positions

2.3 经验要求

对本科学历人才的经验要求,大多在3~4年,其次是在校生/应届生,如图4所示。结果显示,对在校生/应届生的需求岗位最多的是:研发方向(数据研发、研发工程师)、软件开发;其次是数据分析、数据工程师。要求一年经验最多是:数据分析、数据开发;要求2~4年经验的岗位更多的是:数据开发、开发工程师、数据分析师和工程师。要求5~7年经验以上的岗位更多的是架构方向、开发方向。随着经验年限要求越高,薪资水平也越高。

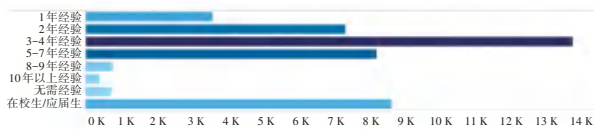


图4 不同经验要求的人才需求条形图

Fig. 4 Bar chart of talent demand for different experience requirements

2.4 地域分布

全国主要城市均需求大数据人才,如图5所示,需求量最大的城市依次是上海、北京、深圳、广州、杭州、成都,这六大城市的需求占总需求的57.6%。除此之外南京、武汉、合肥、西安、珠海、济南等地需求也比较大。可见,这些城市是数字经济发展沃土,是人才聚集地。

2.5 公司行业及性质需求

公司行业字段有多个行业标签,有些公司跨不同行业,本研究默认以第一个作为该公司的行业标签。人才需求行业主要来源于计算机软件、互联网/电子商务、通信/电信/网络设备,金融/投资/证券,

计算机服务(系统、数据服务、维修),电子技术/半导体/继承电路,专业服务(咨询、人力资源、财会)。本文筛选出招聘总人数在平均值之上的16个行业,绘制出行业需求饼图,如图6所示。



图5 大数据人才工作地点分布地图

Fig. 5 Distribution map of big data talents' work places

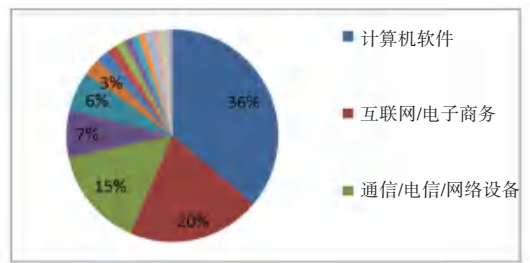


图6 行业需求饼图

Fig. 6 Pie chart of industry demand

此外,民营公司对大数据人才的需求最大,占到一半。然后依次是国企(20%)和上市公司(13%)。

2.6 技能需求和人才素养需求

职位详情字段中,用人单位描述了对招聘人才的能力、素养等要求。采用文本挖掘方法,利用jieba分词工具,对每一条记录的职位详情去除停用词以后进行分词,然后进行词频统计,如图7所示。排在前列的软件和技术要求有PYTHON、JAVA、HADOOP、SPARK、excel、spss、mysql等,这些是大数据应用过程中所需要的工具和平台,用来对数据进行采集、存储、计算、分析等。对人才素养的要求更强调团队合作、良好的沟通能力。

若按照岗位名称分类查看工作要求,需求量最大数据分析岗位的描述词云,如图8所示。可见数据分析岗位要求熟悉python、R语言、excel或spss等热门编程语言或工具;掌握sql数据库;具有数学、统计学、数据挖掘等理论知识;或者应用相关技术,通过数据揭示业务问题。

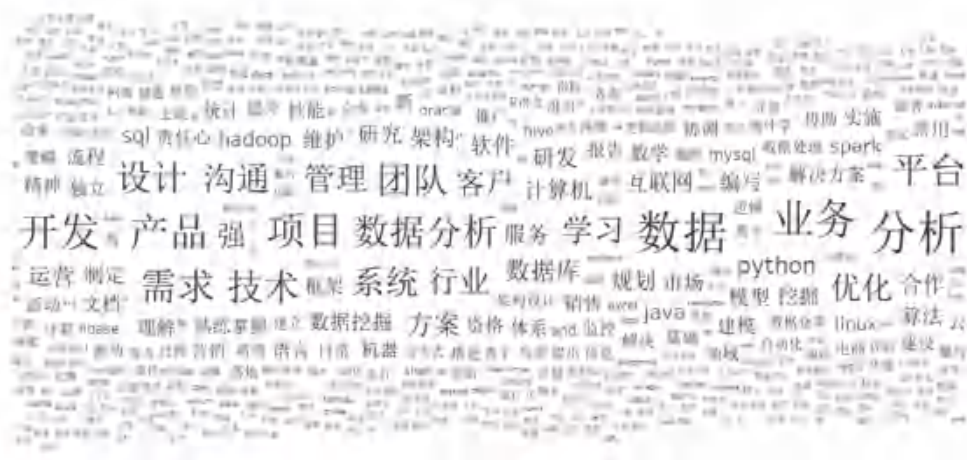


图 7 职位详情描述词频统计云图

Fig. 7 Word cloud chart of job description frequency statistics



图 8 数据分析方向的职位详情描述词频统计云图

Fig. 8 Word cloud chart of job description frequency statistics in data analysis direction

此外,根据需要可以得到不同条件下职位描述词云图,在此则不一一赘述。

3 结束语

经过研究和分析大型招聘网站的招聘数据,获知当前市场对大数据人才的需求是多行业、多层次、多分类的,薪资较高、需求量大、工作地大多在北上广深、杭州及成都等主要城市。本科高校在设计人才培养方案时,以国家战略、产业行业发展的顶层设计为出发点,及时根据人才市场需求,规划和调整教育教学体系,培养适应多元化发展的创新性人才。在此提出以下建议供参考:

位的需求也不尽相同,高校在“全科教育”的基础上培养学生的某个领域特长,应是较合理的培养方向 and 理念。

(3)专业教育与创新教育互为补充和支撑,全面提升大数据相关专业学生的素质,适应快速发展的社会。

(4)特色培养。每个学校的办学基础不一,有不同的硬软件环境。地方本科院校应进一步分析本地及辐射地区的人才需求,为各地数字经济和新型智慧城市建设项目输送高质量的大数据人才。而跨学科的专业,可深入分析相关行业的特殊需求,培养兼具行业背景和大数据技能的复合型人才。

参考文献

[1] 赛迪网. 2019-2021 年中国大数据市场预测与展望数据[EB/OL]. <https://cloud.tencent.com/developer/news/490716>.

[2] 孙会峰. 2019 中国大数据产业发展白皮书[J]. 互联网经济, 2019(22):10-21.

(1)考虑学生的职业发展路径,参照上述热门岗位经验要求分析,就业 5 年内的职位发展路径有迹可循,依次是数据研发、数据分析、开发和运营、架构。

(2)有针对性的培养学生技能。不同行业对岗