

中国人民大学经济学院以研究生毕业同等学力

申请硕士学位论文写作报告

姓 名： 钱悦

资格证号： 71041166

专业名称： 企业经济学

拟定学位论文题目： 基于文本挖掘技术的

中小企业会计人才技能需求研究

报告日期： 2022年1月22日

一、选题依据

|  |
| --- |
| 1.目的及意义（800字以内）（主要内容：阐述选题要解决什么问题，选题有何理论和现实意义）  大数据应用已渗透到企业财务领域，会计专业呈现信息化和管理化趋势，一方面由于国家推行的金税工程及后续配套政策，小微企业更倾向于选择成本更低的代理记账模式，代理记账公司则通过信息化整合系统集中处理小企业记账问题，缩减人力成本；而大型公司通过建设财务共享中心将会计工作电子化、模块化，进一步缩减人力成本、提供管理分析数据。  随着疫情对中国经济特别是小微企业运营的进一步冲击，社会对于会计基础性操作岗位的需求进一步缩减；即使忽略疫情的影响，大学也有取消会计相关专业基础学科设置的趋势，会计专业和会计基础职位的就业前景愈不明朗。面对日趋严峻的就业形势，对财会专业设置的未来构想，除了需结合本专业的实际情况，也需要密切关注计算机和大数据行业的发展趋势，对于已被机器取代或预计未来会被取代的技能进行更新，并加入实用性更强的技能。  随着区域经济发展，产业结构的调整带来劳动力就业结构的改变，经济科技的迅猛发展使生产方式由劳动密集型转向技术密集型、进而转向知识密集型，对劳动者素质的要求也越来越高，企业对技能型人才的需求日益增长。当前我国产业结构中技术型产业所占的比重越来越大，人才结构与内涵要求也随之变化，如何实时、迅速感应市场对人才的需求，对人才培养、人才结构调整有重要意义。  本文研究现阶段北京市场上会计岗位的人才需求是十分有意义的，招聘网站发布的会计岗位招聘信息是最能够反映市场对会计人才需求的数据。分析市场需求，对有意从事企业会计岗位的求职者们起到求职引导作用，使得他们了解目前最新的会计岗位市场需求动态，譬如该岗位对哪些技能的要求较高，从而查漏补缺，向市场需求的方向努力靠拢；也是人才输送单位进一步根据市场调整企业人才供给的重要手段之一，使得人才供给和人才结构趋向合理向市场需求的方向努力靠拢，日趋合理。 |

|  |
| --- |
| 2.文献综述（3000字左右）（主要内容：做文献梳理和研究动态的综述，归纳已有的研究所做的工作，形成了哪些共识？列举出在哪些问题上仍未形成共识？各种不同的观点是什么？针对目前的研究，你发现了哪些问题想要继续研究？）  本文以职业学校财经商贸类专业下设的会计专业为例，通过研究网络公开的招聘网站数据，分析该专业所需的人才技能。所要参考的前人研究成果主要涉及会计类专业人才所需技能发展的研究、人才技能的数据来源及方式方法以及文本挖掘具体方法的应用等方面。  2.1 企业会计人才的技能需求发展  会计专业是财经商贸类专业的下设专业。财经商贸类专业是中等职业教育专业最大的专业类别之一，在现阶段存在教学理论与社会应用实践不匹配的问题，有大量学者将研究的目光投向财经商贸类专业。严瑾玉（2013）[6]分析了该专业的教学现状及存在问题，指出教学课程与社会实际需要的应用能力存在偏差，要加强财经商贸专业改革，解决学生就业问题。林鹏（2016）[7]以福建省为研究区域，关注职业教育中的实践教学，以财经商贸专业中的课堂实践教学、实训实践教学、引入企业综合项目实践教学、企业跟岗实践教学、企业顶岗实践教学等具体教学实践形式为研究对象，从中发现实践教学中存在的问题，为优化人才培养模式提出建设性意见。武紫云（2018）[9]以学生的核心职业能力为研究目标，紧扣职业教育“以就业为导向”的号召，从培养现状的基础上，探讨中职学校财经商贸学生应该具备的职业核心能力，解析培养过程中遇到的问题，构建了课程体系，促进职业核心能力的提升，并提高学生的就业竞争力。  2.2 人才需求分析的数据来源  不同行业和企业对人才的需求标准和关注点各有不同，这一直是一个广受国内外专业学者研究的课题，因此学术界存在着不同专业的学者通过各种方法和途径研究其所在专业的人才需求特征的领域研究。通过阅读相关文献，总结发现目前通过获得人才需求分析原始数据的方式主要为问卷调查法和分析招聘网站公开信息。而招聘网站获取行业人才的需求数据的研究过程可提炼为：确定搜索范围（时间范围、设置条件及筛选标准等），获取招聘网站基础数据，对基础数据进行二次筛选（去除无效数据），处理招聘数据信息。  在中等职业教育实践领域，获得人才需求的数据主要通过对校企合作的相关企业的访谈和问卷调查后整理相关基础数据，确定技能和需求及专业设置后再通过由校内专业教师、企业专家和已经工作的学生代表等组成的专业评估委员会进行评估。这种方法由于企业数量和自身视野的限制，对人才的需求认识比较片面，而海量的公开招聘数据无疑是更好的人才需求原始数据。  在获取招聘网站基础数据过程中，分为人工获取和使用大数据技术手段及工具爬取招聘网站数据两种方式。方式的选择主要基于在招聘网站搜索后的结果数量。当数据在上百条时，研究者倾向于选择人工方式进行保存数据和筛选数据，例如潘香笑和丁敬达（2017）[8]在企业信息咨询职业核心技能需求研究中，使用人工方式获取了297条数据；罗玉波和张东霞（2018）[10]在其对会计人才需求问题研究中，使用人工保存的方式获取了501份招聘广告。  当数据更多，达到上万条时，研究者倾向于选择爬取招聘网站数据的方式，例如，宋齐明（2018）[11]通过网络爬虫工具经去重共爬取了15653条网络招聘信息。经过在招聘网站的初步搜索，获取了大量数据，爬虫手段能够处理的数据量更大、更客观真实的反映招聘信息，且对于后续工作具有可复制性和参考意义，得到的结论也更具代表性。  在处理招聘数据信息过程中，信息处理方式也分为人工处理和自动化处理。人工处理数据的方式广泛存在于社科类及文科的行业研究中，自动化处理方式则主要出现于数据分析、IT等技术类行业的人才需求分析中。罗玉波、张冬霞（2018）[10]从各大招聘网站共收集了财务会计人员招聘广告501份，进行人工筛选后形成有效样本498份，将会计人才需求分类为特殊要求、一般要求和基本要求，通过词频探究会计人才市场的需求；潘香笑、丁敬达（2017）[8]通过人工手段对274条招聘广告进行了筛选和整理，从技术维度、商务维度和综合维度将数据构成核心技能的类目。但人工处理数据在准确性、客观性上存在一定的偏差，并且需要付出大量的人力和时间成本。Litecky,C.（2010）[5]等通过自动化处理并进行文本分析提取计算机工作的技能术语的文本，采用层次聚类和K-Means聚类方法，将IT岗位类型分成了20类，并给出了不同类型岗位的技能需求组合。詹翠芬（2018）[12]也通过K-Means聚类划分了互联网行业的岗位类型，再对各岗位类型进行相关性分析。本文使用自动化处理的方法处理海量招聘数据。  2.3文本挖掘技术方法研究  关于招聘信息方面的研究，国外学者运用文本挖掘方法研究人才市场需求已经形成比较明确的体系。Todd（1995）[1]等人在报纸上收集了美国和加拿大的信息系统岗位信息，提取关键词并进行频数分析；Lee（2006）[2]在对信息技术管理岗位进行分析时，构建了技能分类目录，产生了词典的雏形；Sodhi,M.S.和 Son,B.G.（2010）[3]结合以上研究，从招聘网站采集了运筹学专业的就业信息，构建了技能词典和关键词词典，在进行频数分析的基础上又进一步做了交叉分析和相关性分析，研究了运筹类专业技能在不同行业需求的差异；Smith，D.和Ali，A.（2014）[5]等人从招聘网站收集程序员的招聘信息，利用关键词索引技术，分析了近几年编程语言的市场行情，为计算机专业的课程设置提供参考意见。这些国外研究说明利用文本分析的方法对不规则数据进行研究是可行的。 |

二、研究方案

|  |
| --- |
| 1.论证方法及数据来源（主要内容：说明论证拟采用的方法，如数理模型法、计量分析法等等，以及需要用到的数据及其来源）  根据猎聘网网络招聘数据中的工作地点、薪资分布、学历要求、经验要求、公司性质、公司规模、岗位职责和岗位要求等信息，基于 Excel、Python等工具，利用他们处理数据的能力对以上内容进行分析和挖掘，具体步骤如下：  第一步：数据获取：使用 Python编程语言中的Scrapy网络爬虫框架爬取招聘网站数据、收集原始信息，作为数据爬取工具进行数据整理工作。  Python是一款功能强大的免费编程语言，因其学习方便、功能强大、可调用的模块多，常用于数据分析及人工智能领域。Scrapy库是Python中的专门模块。Scrapy框架的主要原理为：  1、爬虫引擎（Scrapy Engine）：负责控制数据流在系统中所有组件中流动，并在相应动作发生时触发事件。  2、调度器(Scheduler)：从引擎接受请求了爬取网站数据、收集信息的框架并控制请求顺序，以便之后引擎请求时提供给引擎。  3、下载器(Downloader)：负责获取页面数据并提供给引擎，而后提供给蜘蛛（Spider）。  4、蜘蛛（Spiders）：用于访问特定网址、分析response并提主题内容。  5、项目管道（Item Pipeline）负责处理被spider提取出来的主题。典型的处理有清理、 验证及持久化(例如存取到数据库中)。  6、下载器中间件(Downloader middlewares)：连接框架引擎与下载器。  7、Spider中间件(Spider middlewares)：处理spider的输入(response)和输出(items及requests)等解析部分。  8、数据流(Data flow)：Scrapy中的数据流由执行引擎控制。  a6285dff405e1b6be04b9f4378fed7c具体实现过程如图1[13]所示：  图1  第二步：数据清洗：整合采集到的招聘信息，对岗位信息进行筛选、去重、整理等处理，剔除部分不良样本和无效样本。  第三步，对经处理的招聘信息进行统计分析。按照岗位所在地（具体到北京区县）、公司类型、公司规模、学历、工作经验、薪资等进行岗位统计分析。  第四步，数据预处理：主要包括文本分词、停用词过滤和构建岗位信息的技能词典。利用 Python软件对会计岗位的岗位职责和岗位内容进行分词处理，建立LDA主题模型，得出主题和关键词，构建大数据类各岗位的职位画像，得出企业技能需求。 |
| 2.核心观点（主要内容：初步阐述可能得到的观点及结论）  面对日趋严峻的就业形势，会计专业的未来构想，除了需结合本专业的实际情况，也需要密切关注计算机和大数据行业的发展趋势，对于已被机器取代或预计未来会被取代的技能进行更新，并加入实用性更强的技能。  通过分析公开的网络招聘信息分析企业人才技能需求，从招聘公司情况、所需人才学历、人员经验和内容等方面，分析人才技能需求。  在该选题下，本文试图通过分析网络招聘广告研究以下问题：   1. 通过数据分析等定量研究的提供指导性信息。 2. 通过文本挖掘技术运用自动化手段获取智联招聘及猎聘网的公开招聘信息，进而对原始数据进行文本分词等预处理并利用LDA（Latent Dirichlet Allocation）主题模型进行归类整理，实现财会岗位技能分析流程的全自动化流程及应用场景。   第三，获得人才技能需求数据，包括技能类别以及具体要求，明确企业会计人才的需求和发展方向。 |
| 3.创新之处（主要内容：简要阐述创新点，比如方法创新、方向创新、观点创新等等）  方法创新：利用技术手段挖掘人才储备，将文本挖掘技术和数据分析方法运用于企业会计人才需求研究领域，实现实时更新、可复制流程的人才需求分析方法。  方向创新：真正从市场公开信息的角度理解会计人才需求变化和发展情况，综合市场数据获得真实结果。 |

|  |
| --- |
| 4.参考文献（顺序和格式参考模板如下，参考文献应当主要是近5年的相关资料，填写时，删掉以下参考模板）  参考文献  [1] Todd,P.A., Mckeen,J.D., Gallupe R.B. The evolution of IS job skills: a content analysis of IS job advertisements from 1970 to 1990 [J ] Mis Quarterly, 1995,19(1): pp.1-27.  [2] Lee, S.M., Lee,C.K. IT managers' requisite skills [J]. Communications of the Acm, 2006,49(4):pp.111-114.  [3] Sodhi,M. S. and Son.,B.G. Content analysis of OR job advertisements to infer required skills [J]. The Journal of the Operational Research Society, 2010,61(9): pp.1315-1327.  [4]Litecky,C., Aken A., Ahmad A. Mining for Computing Jobs[J]. IEEE Software,2010,27(1):pp.78-85.  [5]Smith, D.,Ali,A. Analyzing Computer Programming Job Trend Using Web Data Mining [J]. Issues in Informing Science and Information Technology, 2014.  [6]严瑾玉.浅析财经商贸专业创新式发展[J].现代营销（学苑版）,2013(9):52.  [7]林鹏.浅析福建省中职财经商贸类专业实践教学[J].价值工程,2016,35(31):189-191.  [8]潘香笑,丁敬达.企业信息咨询职业核心技能需求研究——基于招聘广告的内容分析[J].情报理论与实践,2017,40(02):56-59.  [9]武紫云. 中职学校财经商贸类专业学生职业核心能力培养研究[D].天津职业技术师范大学,2018.  [10]罗玉波,张冬霞.会计人才需求问题研究——基于会计人才的市场调查[J].广东石油化工学院学报,2018,28(02):56-60.  [11]宋齐明.劳动力市场需要什么样的本科毕业生——基于近1.4万条招聘信息的量化分析[J].中国高教研究,2018(03):56-60.  [12]詹翠芬.基于数据挖掘的互联网行业岗位类型分析[J].计算机产品与流通,2018(07):136-138.  [13]https://blog.csdn.net/baidu\_32542573/article/details/79415947 |

|  |
| --- |
| 5.论文提纲（写到二级标题）  题 目： 基于文本挖掘技术的中小企业会计人才技能需求研究  主题词： 文本挖掘，中小企业，会计  摘要  abstract  引言  第1章 绪论  1.1 研究背景及意义  1.2 研究方法与思路  1.3 文章结构  1.4 创新点  1.5 数据采集与数据预处理   1. 文献综述   2.1企业会计人才的技能需求发展  2.2人才需求分析的数据来源  2.3文本挖掘技术方法研究   1. 企业会计岗位的理论分析   3.1 企业会计岗位的主要特征  3.2 各地区会计岗位需求主题词提取  3.3 文本聚类分析理论  第4章 企业会计岗位的实证分析  4.1 基于LDA模型的主题词提取  4.2 北京市企业会计岗位对比分析  第5章 结论与展望  5.1 结论与展望  5.2 不足与建议  参考文献  附录  致谢 |