

第十一届“泰迪杯” 数据挖掘挑战赛

优秀 作品

作品名称：泰迪内推平台招聘与求职双向推荐系统构建

荣获奖项：一等奖并获网宿创新奖

作品单位：华南师范大学

作品成员：刘晨 姚詠滋

封面为后期添加，原图没有此页。

泰迪内推平台招聘与求职双向推荐系统构建

摘要

为了缓解招聘流程中招聘方和求职方信息不对称的问题，更好地为双方进行推荐。本文建立了基于贪心算法和多目标规划的招聘求职双边匹配模型，获取使履约率最高的推荐方案。

对于问题一，通过 API 对招聘信息和求职信息数据进行爬取，并通过对 json 数据的分析对数据进行分类整理，最后对爬取的数据进行初步数据预处理以及数据描述。

对于问题二，使用爬取的招聘信息和求职信息数据以及 Python 的 plotly 库来构建画像：根据采集的企业招聘信息，从招聘岗位、学历要求、公司行业、公司类型、薪资待遇、企业工作地点、员工数量和岗位经验等多个方向建立招聘信息画像；根据采集求职者求职信息，从预期岗位、期望行业、薪资需求、知识储备、学历和性别等多个方向建立求职者画像。最后根据建立的画像给求职者提供建议。

对于问题三，由于招聘求职双方对彼此都有自己的期望与要求，因此在进行推荐之前必须了解求职者与岗位的匹配度和求职者对岗位的满意度。本文建立了基于向量相似度的岗位匹配度和求职者满意度模型。首先选取了 6 个重要指标，以职位和求职者的需求作为标准向量，对双方的指标信息进行量化处理，生成招聘信息向量和求职者信息向量，采用余弦相似度的计算公式求出其与标准向量的相似度，得出岗位匹配度或求职者满意度。

对于问题四，根据招聘流程，为了使招聘双方适配且履约率达到最高，本文从两个角度出发，分别建立了基于贪心算法和基于多目标规划的招聘求职双边匹配推荐模型，最后根据模型结果选取履约率最高的模型。从贪心算法的模型出发，本文将每一轮招聘看做是一个子问题，通过对职位和求职者的选择策略进行优化，做出每一轮最优的选择，以获得全局最优的方案；从多目标规划的模型出发，本文以招聘双方的感知效用最大化和签约人数最大化为目标，通过线性加权的方式将多目标模型转化为单目标模型并进行求解。比较两种模型得出最终的人岗匹配方案。

关键词：招聘与求职画像；人岗匹配；向量相似度；贪心算法；多目标规划

目 录

摘 要	I
目 录	II
第一章 引言	1
1.1 赛题背景	1
1.2 问题重述	1
1.3 问题分析	2
1.3.1 问题一：招聘信息爬取	2
1.3.2 问题二：招聘与求职信息分析	2
1.3.3 问题三：构建岗位匹配度和求职者满意度模型	3
1.3.4 问题四：构建招聘求职双向推荐模型	3
第二章 数据分析处理	3
2.1 爬虫获取数据集	3
2.1.1 获取 API	3
2.1.2 分析 json 数据	5
2.1.3 异常处理	5
2.1.4 初步数据预处理	6
2.2 数据描述	6
2.3 招聘信息与求职信息画像	7
2.3.1 招聘信息画像	7
2.3.2 求职信息画像	13
第三章 基于向量相似度的岗位匹配度和求职者满意度模型	16
3.1 模型假设	16
3.2 指标类型	16
3.3 符号说明	17
3.4 岗位匹配度模型	18
3.4.1 招聘需求指标	18
3.4.2 招聘需求指标的量化方法	19
3.4.3 相似度计算	20
3.4.4 基于向量相似度的岗位匹配度模型	21
3.5 求职者满意度模型	22
3.5.1 求职需求指标	22
3.5.2 求职需求指标的量化方法	22
3.5.3 相似度计算	24

3.5.4 基于向量相似度的求职者满意度模型	25
3.6 模型结果	25
第四章 招聘求职双边匹配推荐系统	26
4.1 算法选择	26
4.2 符号说明	26
4.3 算法理论基础	27
4.3.1 贪心算法	27
4.3.2 多目标规划	27
4.3.3 整数规划	28
4.4 基于贪心算法与策略优化的招聘求职双边匹配模型	28
4.4.1 求职者选择策略	28
4.4.2 岗位选择策略	29
4.4.3 招聘求职双边匹配推荐模型	30
4.5 基于多目标规划的招聘求职双边匹配推荐模型	31
4.5.1 感知效用矩阵构建	32
4.5.2 决策模型构建	32
4.5.3 决策模型求解	33
4.5.4 决策步骤整理	33
4.6 模型结果	34
第五章 总结与展望	34
5.1 总结	34
5.2 优缺点分析	34
5.2.1 优势	34
5.2.2 不足	35
5.3 展望与优化	35
参考文献	36

第一章 引言

1.1 赛题背景

在新时代背景下，随着大学生毕业人数屡创新高，就业需求也随之激增，大学生求职问题形势愈发严峻，已成为广泛关注的社会热点。求职就业问题呈现出“两头难”现象：一方面毕业生“找工作难”，对企业和自身定位缺乏明确了解与认知，盲目海投难以找到合适的工作；另一方面企业“招工难”，存在不少职位无人问津的现象。归根结底是因为在求职招聘过程中，招聘方与求职者的信息不对称。[1]

此外，受疫情影响，线下招聘受到限制，加之场次有限的宣讲会、招聘会满足难以招聘方与求职者的需求，因此，跨越地域、实时共享的网络招聘应运而生。许多企业的招聘目前都改为线上进行，避免了时间和空间的限制。同时，招聘需求不断上涨，近六成企业招聘需求增加，其中需求量较大的科技研发、数字化、蓝领技能岗位都存在不同程度的人才短缺。[1][3]

针对招聘求职问题，泰迪内推平台——聚焦于“大数据+”和“人工智能”领域的求职招聘网站，融合了多家企业发布的招聘信息，也为求职者提供求职信息的展示。而为了缓解毕业生就业压力，同时满足企业对人才的需求，泰迪内推平台会定期为高校学生提供优质岗位推荐，解决毕业生就业的同时也缓解企业用人难的问题，为校企之间搭建起资源互换的桥梁，力求实现人才的供需对接和教育资源转化，通过深化产教融合，促进教育链、人才链、产业链与创新链有机衔接。

因此，对招聘信息进行分析研究，了解不同职业领域的需求特点，挖掘兴起的数据类行业相应的人才需求现状及发展趋势，为广大求职者提供正确的就业指导不仅有重要具有重大的实用价值，也具有重要理论意义。

1.2 问题重述

1. 问题一：招聘信息爬取

从泰迪内推平台的“找工作”页面和“找人才”页面，爬取所有招聘与求职信息并进行整理。

2. 问题二：招聘与求职信息分析

构建招聘与求职画像：根据采集的企业招聘信息，从招聘岗位、学历要求、岗位需求量、公司类型、薪资待遇、岗位技能、企业工作地点等多个方向建立招聘信息画像；根据采集求职者求职信息，从预期岗位、薪资需求、知识储备、学历、工作经验等多个方向建立求职者画像。

3. 问题三：构建岗位匹配度和求职者满意度模型

根据招聘信息与求职者信息，构建岗位匹配度和求职者满意度的模型，基于该模型，为每条招聘信息提供岗位匹配度非 0 的求职者，将结果进行降序排序；为每位求职者提供求职者满意度非 0 的招聘信息，将结果进行降序排序。

4. 问题四：构建招聘求职双向推荐模型

假设招聘流程如下：设某岗位拟聘 n 人，泰迪内推平台向企业推荐岗位匹配度非 0 的 n 位求职者发出第一轮 offer，求职者如果收到多于 1 个岗位的 offer，则求职者选取满意度最高的岗位签约，每个求职者只允许选择 1 个岗位签约。第一轮结束后，平台根据当前各招聘信息的剩余岗位数，向后续被推荐求职者发出第二轮 offer，如此继续，直到招聘人数已满或者向所有拟推荐求职者均已发出 offer 为止。

履约率高低是评价平台的推荐系统优劣的重要指标。本文的履约率定义为：

履约率=所有岗位的签约人数之和/所有拟聘岗位人数之和。

按照上述招聘流程为平台设计一个招聘求职双向推荐模型，使得履约率指标达到最高，并根据模型给出最终的招聘岗位与求职者签约成功的结果。

1.3 问题分析

1.3.1 问题一：招聘信息爬取

问题一要求从泰迪内推平台上爬取所有招聘与求职信息并整理好，是一个典型的数据采集和整理问题。可以通过 API 进行爬取、并对获取的 json 数据进行分析整理和数据预处理得到招聘信息和求职信息数据。

1.3.2 问题二：招聘与求职信息分析

问题二要求根据采集的招聘信息和求职信息，从多方面建立招聘信息画像和求职者画像，并进行分析。先对数据进行清洗、去重和异常值、缺失值处理等预处理操作；然后对数据进行归类整理，按照不同分类方式进行分析。在建立招聘信息画像时，对招聘信息情况（如招聘岗位、薪资待遇）、企业各方面情况（如公司类型、公司行业、员工数量和企业工作地点等）与要求（岗位经验和学历要求等）按照不同类别整理出来，绘制相关的图像，直观清晰地了解招聘信息；在建立求职者信息画像时，对求职者各方面情况（学历层次和知识储备等）与需求（预期岗位、期望行业、薪资需求等）按照不同类别整理出来，绘制相关的图像，直观清晰地了解求职信息。

通过建立并分析招聘信息和求职者画像，可以了解市场需求的趋势和变化，为企业招聘人才提出改进建议，也为求职者提供更好的求职建议与方向。

1.3.3 问题三：构建岗位匹配度和求职者满意度模型

问题三要求根据数据建立岗位匹配度和求职者满意度模型，类似双方对彼此的评价。因此我们可以通过按照岗位和求职者的需求（岗位、行业、薪酬和学历等），对二者信息的每个指标进行量化，以二者的各自的需求作为标准向量，应用向量相似度来计算，通过归一化处理，得出岗位匹配度与求职者满意度。我们可以使用矩阵来表示岗位和求职者之间的匹配度和满意度。矩阵中每一个元素表示一个岗位和一个求职者之间的匹配度或满意度。其中，由于每一个岗位或求职者都有最低的基本要求，若不满足，则匹配度或满意度为 0。

1.3.4 问题四：构建招聘求职双向推荐模型

问题四要求为平台设计一个招聘求职双向推荐模型，使得系统为企业推荐的人才情况能使招聘的履约率达到最高。由于这是一个双向的模型，需要充分考虑岗位和求职者的想法，实际上就可以将其看做是人岗匹配问题，关键在于人才推荐的决策和对双方的充分考虑，因此在正式建立相关模型时，先采用问题三中建立的模型，来计算出岗位匹配度和求职者满意度。

在此基础上，从两个方面出发设计推荐算法。一方面，结合招聘流程企业和求职者的选择行为，对其策略进行分析制定，应用贪心算法，为企业推荐求职者及帮助求职者选择，通过 python 求解出最终的人岗匹配结果。另一方面，考虑到推荐模型的目的是推荐双方都满意的求职者，且能够使得履约率尽可能高，本文也建立了多目标规划模型，尝试通过 lingo 和 MATLAB 求解得到最优的人岗匹配结果。

第二章 数据分析处理

2.1 爬虫获取数据集

2.1.1 获取 API

本文爬虫是通过 API 接口进行的，API 的主要目的是提供应用程序与开发人员以访问一组例程的能力，而又无需访问源码，或理解内部工作机制的细节。图 1 和图 2 为获取对应 API 地址的示例，其中图 1 为找人才界面的 API 地址，图 2 为获取人才详细信息界面的 API 地址。类似的可以获取找工作界面和工作详细信息界面的 API 地址。

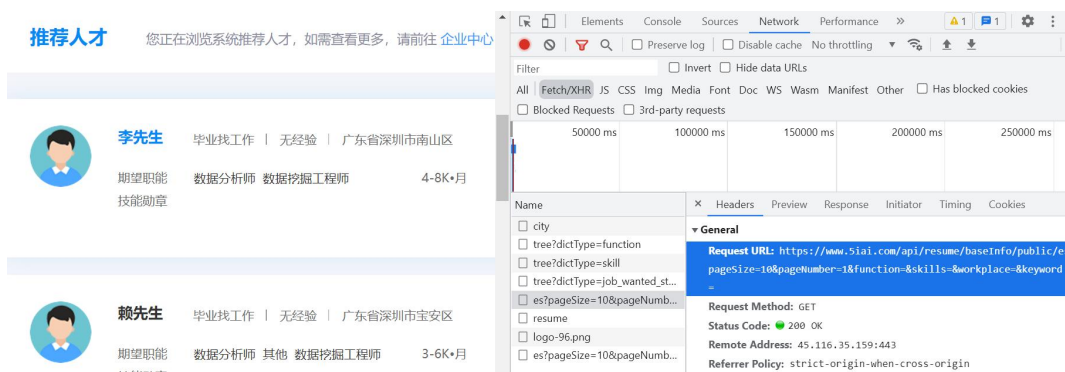


图 1 获取找人才界面 API 地址



图 2 获取人才详细信息界面 API 地址

网页是动态的，数据在不断更新，为方便后续数据的分析与应用，本文采取于 4 月 18 日爬取的静态数据进行研究，获取的 API 地址如下：

(1) 找工作界面的 API 地址

“ <https://www.5iai.com/API/enterprise/job/public/es?pageSize=10&pageNumber={}&willNature=&function=&wageList=%5B%5D&workplace=&keyword=>”

其中 pageNumber 为页码，总共 159 页；每一页有 10 个项目；总共 1581 个项目。

(2) 工作详细信息界面的 API 地址

“<https://www.5iai.com/API/enterprise/job/public?id=1629824814967554048>”

该网址根据对应 ID 号来进行访问，如上述网址根据 ID 号 1629824814967554048 进行访问。

(3) 找人才界面的 API 地址

“ <https://www.5iai.com/API/resume/baseInfo/public/es?pageSize=10&pageNumber={}&function=&skills=&workplace=&keyword=>”

其中 pageNumber 为页码，总共 1093 页；每一页有 10 个项目；总共 10923 个项目。

(4) 人才详细信息界面的 API 地址

“<https://www.5iai.com/API/resume/baseInfo/public/1579833809761861632>”

该网址根据对应 ID 号来进行访问，如上述网址根据 ID 号 1579833809761861632 进行访问。



图 4 爬虫异常数据示例

最后招聘信息总共 1581 个，爬取的实际数据为 1575 个，异常数据为 6 个；求职信息总共 10923 个，爬取的实际数据为 10877 个，异常数据为 46 个。

2.1.4 初步数据预处理

对爬取的招聘信息数据进行初步的数据预处理，如下：

1.对招聘信息进行数据预处理：

- (1) 根据 ID 号的含义，ID 号不能有重复值，检查发现招聘信息 ID 号没有重复值；
- (2) 删除公司地址中无效值“.”，总共 4 个；
- (3) 将岗位经验中“经验不限”、“不限”和“0”数据集成为“0”。

2.对求职信息进行数据预处理：

- (1) 根据 ID 号的含义，ID 号不能有重复值，检查发现求职信息 ID 号有 2648 个重复值，将重复值删除，剩余 8229 个数据；
- (2) 将预期岗位、期望行业和期望职位中出现的“"[\\\\"不限\\\\"]"”进行标准化，改为["不限"]；
- (3) 去除简历关键词中的无关语言，如“666,777”；
- (4) 去除自我评价中的无关语言，如“qqqq”、“dyyy”；
- (5) 学历中出现多个学历，取最高学历，如“本科 硕士”改为“硕士”。

将初步预处理后的招聘信息保存为“result1-1.csv”文件，初步预处理后的求职信息保存为“result1-2.csv”文件。

2.2 数据描述

爬取的招聘信息数据有招聘信息 ID、企业名称、招聘岗位、员工数、公司类型、

公司行业、公司地址、工作地址省、工作地址市、工作地址区、最低薪资、最高薪资、工作类型、学历、岗位需求量、岗位关键词、技能关键词、岗位经验、职位福利、职位描述，以下是对招聘信息部分数据进行描述：

表 1 招聘信息部分数据描述

数据	数据说明
员工数	有八类，分别为少于 50 人, 50-100 人, 150-500 人, 500-1000 人, 1000-1500 人, 1000-5000 人, 5000-10000 人, 10000 人以上
公司类型	有六类，分别为上市公司、合资、国企、外资、民营、私企
工作地址省、工作地址市、工作地址区	分别指工作地址的省代码、市代码、区代码
工作类型	分为 0、2，对应实习、全职
学历	分为 0、1、2、3、4、5，对应不限，技工、大专、本科、硕士、博士
岗位需求量	0 为不限，其他数字表示具体招聘人数
岗位经验	0 指不限经验

爬取的求职信息数据有求职者 ID、姓名、预期岗位、性别（分为 0、1，对应男、女）、年龄、工作经验、工作性质、期望行业、期望职位、到岗时间、预期最低薪资、预期最高薪资、地区、简历关键词、自我评价、学历、专业、项目经历、比赛经历、语言、技能、证书、工作经历、培训经历。

2.3 招聘信息与求职信息画像

本节使用 2.1 中的招聘信息与求职信息以及 Python 的 plotly 库对招聘信息和求职信息的多个方向进行了分析和可视化，得到了招聘信息画像和求职信息画像，并可以为求职者和企业提供更准确的参考信息。其中招聘信息画像总共从 8 个方向进行了分析，包括招聘岗位、学历要求、公司行业、公司类型、薪资待遇、企业工作地点、员工数量和岗位经验等方面。求职信息画像总共从 5 个方向进行了分析，包括预期岗位、期望行业、薪资需求、知识储备、学历和性别等方面。

2.3.1 招聘信息画像

1. 招聘岗位

首先观察到求职信息数据中的预期岗位共有 9 种类型，分别为“图像处理工程师”、

“大数据开发工程师”、“数据分析师”、“数据挖掘工程师”、“机器学习工程师”、“算法工程师”、“自然语言处理工程师”、“计算机视觉工程师”、“其他”。但招聘信息数据中的招聘岗位数据没有分类，所以需要根据招聘岗位和岗位描述数据来对招聘岗位进行分类。采用关键词匹配的方法进行分类，部分关键词如下表：

表 2 招聘岗位的关键词

招聘岗位	关键词
图像处理工程师	图像处理工程师、图像处理、图像
大数据开发工程师	大数据开发工程师、大数据开发、数据开发、后端
数据分析师	数据分析师、数据分析、数据产品、数据运营
数据挖掘工程师	数据挖掘工程师、数据挖掘
机器学习工程师	机器学习工程师、机器学习、深度学习
自然语言处理工程师	自然语言处理工程师、自然语言处理、自然语言、语义理解、NLP、nlp
计算机视觉工程师	计算机视觉工程师、计算机视觉、视觉、前端
算法工程师	算法工程师、算法

将分类得到的数据保存为“work.csv”文件，统计出各种类型的岗位需求量如表 3：

表 3 各招聘岗位的岗位需求量

预期岗位	岗位需求量	预期岗位	岗位需求量
图像处理工程师	16	算法工程师	73
大数据开发工程师	212	自然语言处理工程师	90
数据分析师	1228	计算机视觉工程师	25
数据挖掘工程师	260	其他	95
机器学习工程师	183		

根据表 3 画出饼图如图,可以看出招聘岗位中数据分析师和数据挖掘工程师这两类需求量较大。

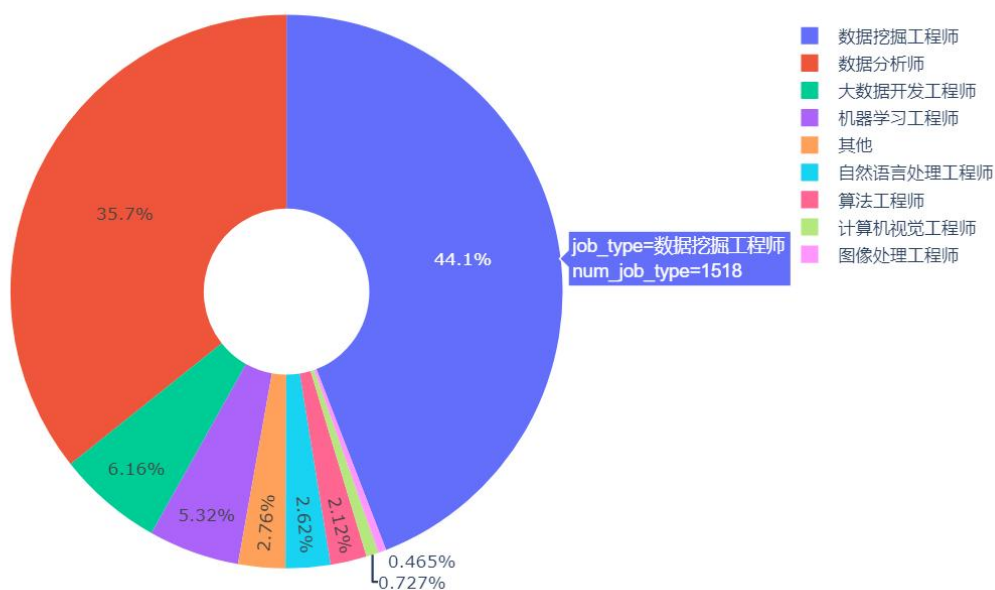


图 5 招聘岗位饼图

2.学历要求

从直方图可以看出,本科学历的需求最大,占比超过四分之一,其次是大专学历。

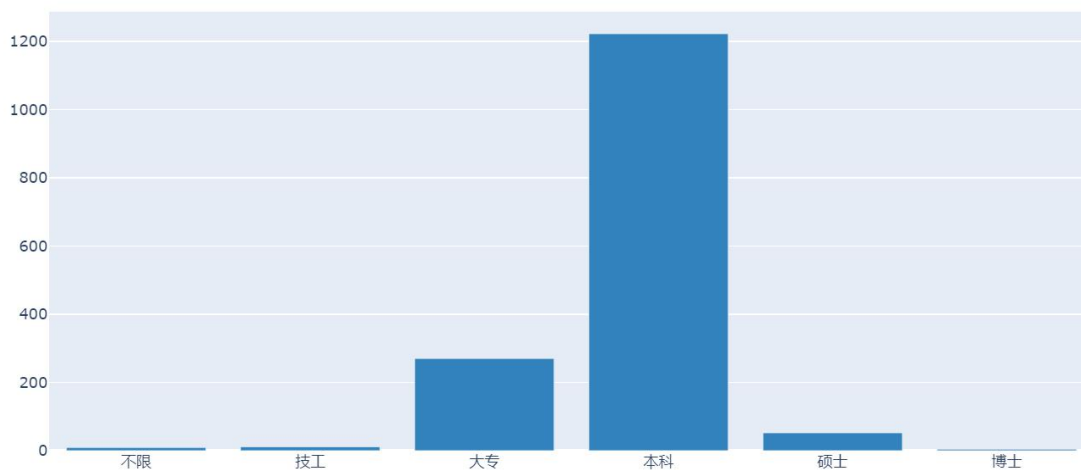


图 6 学历要求直方图

3.公司行业

公司行业共有 34 种，选取发布招聘数量最多的五个行业绘制饼图如图，可以看出互联网行业需求最大，剩余依次是电子商务、软件、大数据、数据服务。

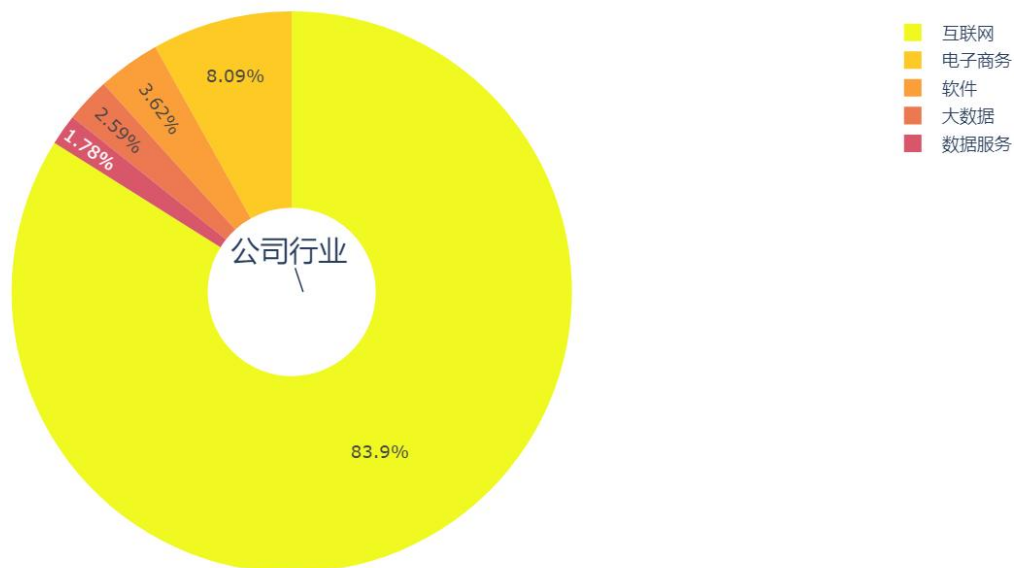


图 7 公司行业饼图

4.公司类型

从饼图可以看出，合资公司的招聘需求最大，占比超过一半，民营公司和上市公司招聘需求比较接近。

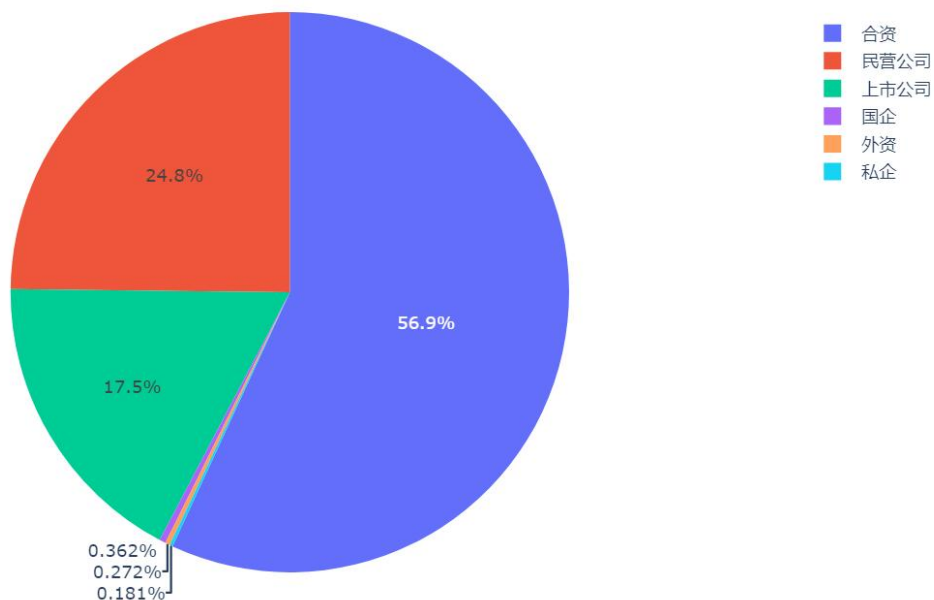


图 8 公司类型饼图

5. 薪资待遇

从箱线图可以看出，最低薪资的范围大概在 2000-2600，最高薪资的范围大概在 2175-4500。

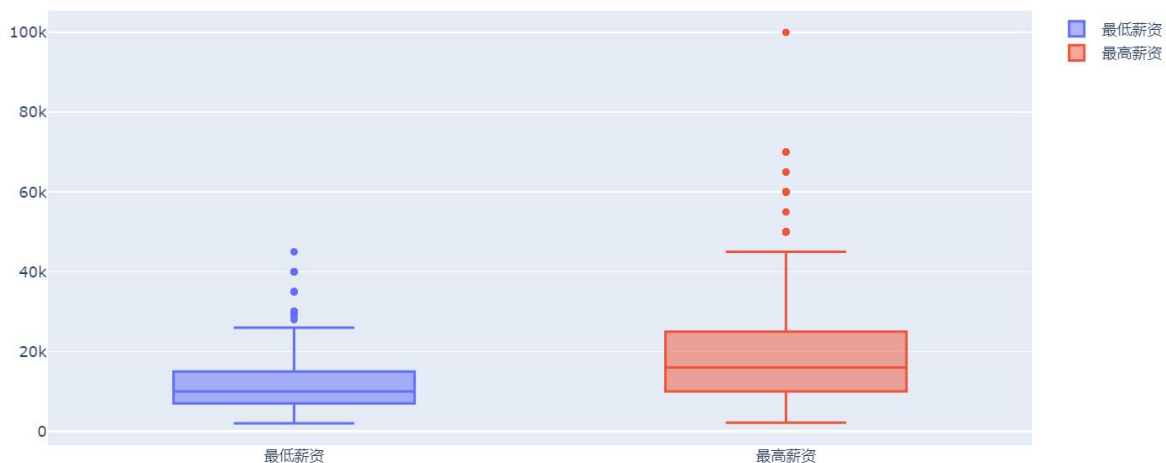


图 9 薪资待遇箱线图

6. 企业工作地点

根据公司地址、工作地址省代码等数据得出企业工作地点。先将工作地址省代码转化为省份，如“440000”的省代码表示广东省。剩余的空白值利用公司地址数据来提取出省份，如“广州大学城(大学城北)创智园 3 栋 4 楼”提取出“广东省”。使用 tableau 得到的地图可视化工作地点的分布如图，可以看出企业工作地点大多数位于广东省。

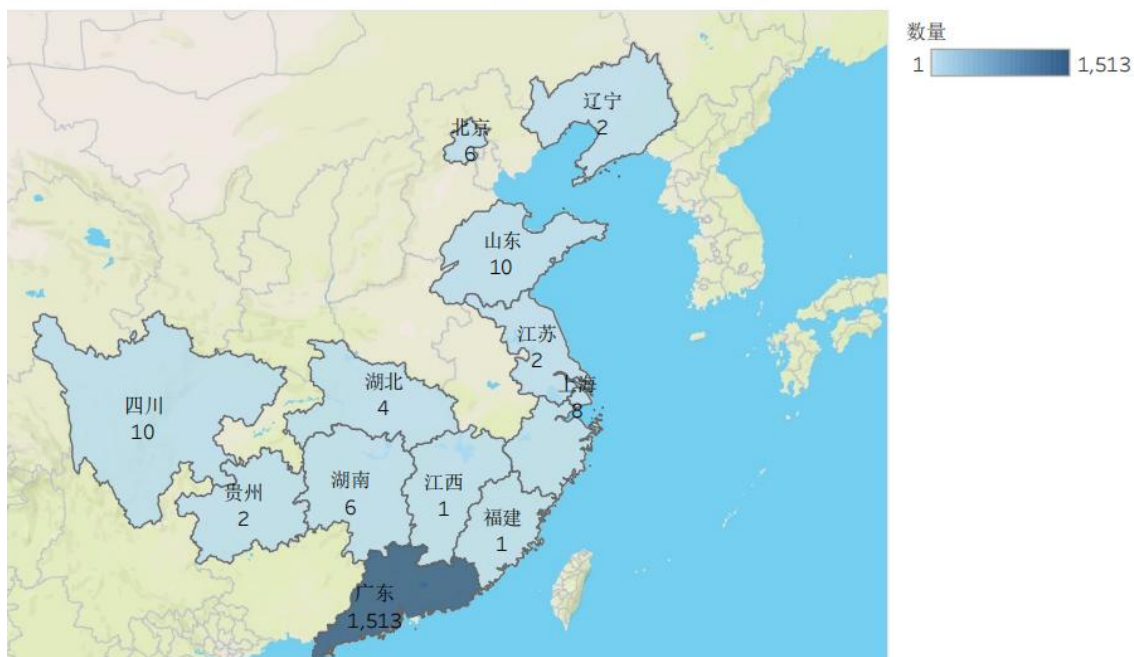


图 10 企业工作地点分布图

7. 员工数量

员工数量共有 8 类，绘制饼图如图，可以看出员工数量为 50-100 人的公司招聘需求量最大。

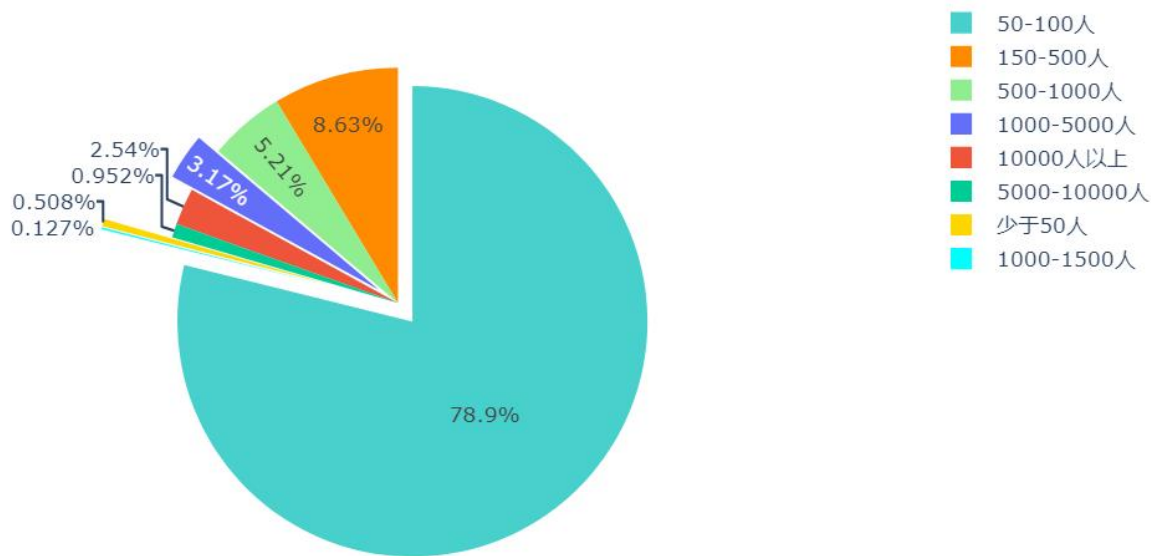


图 11 员工数量饼图

8. 岗位经验

先将岗位经验分为 7 类，分别为“经验不限”、“1 年”、“1-3 年”、“3-5 年”、“5-7 年”、“7-10 年”、“10 年以上”，绘制水平柱状图如图，可以看出绝大多数招聘对于岗位经验没有要求，有一部分公司对岗位经验由 1-5 年的要求。

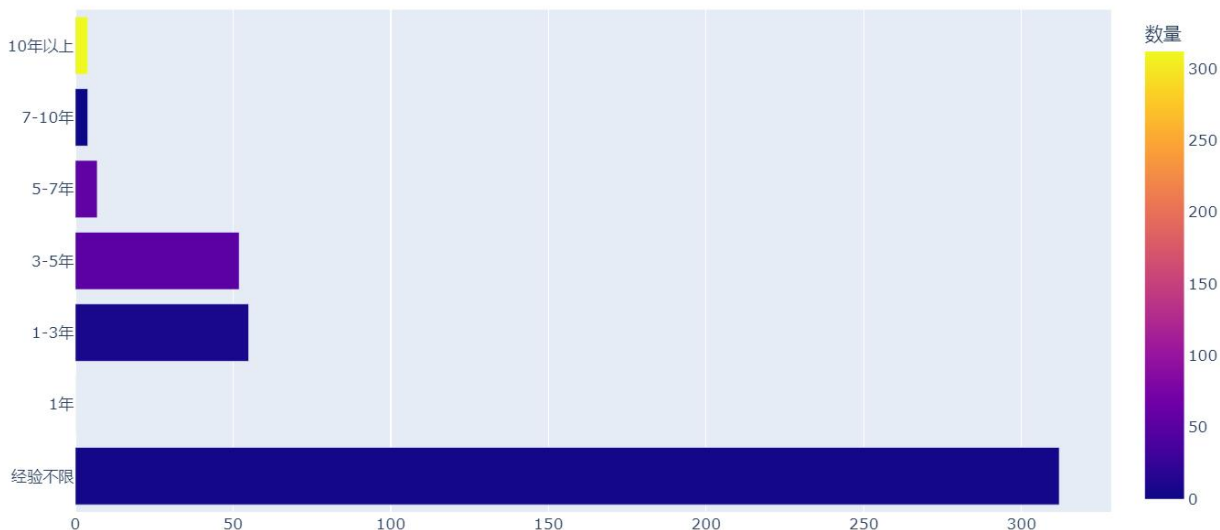


图 12 岗位经验水平柱状图

2.3.2 求职信息画像

1. 预期岗位

预期岗位共有 9 种类型，分别为“图像处理工程师”、“大数据开发工程师”、“数据分析师”、“数据挖掘工程师”、“机器学习工程师”、“算法工程师”、“自然语言处理工程师”、“计算机视觉工程师”、“其他”。先统计出各种类型的岗位需求量如表 4:

表 4 预期岗位的岗位需求量

预期岗位	岗位需求量	预期岗位	岗位需求量
图像处理工程师	11	算法工程师	14
大数据开发工程师	59	自然语言处理工程师	18
数据分析师	8159	计算机视觉工程师	4
数据挖掘工程师	8096	其他	43
机器学习工程师	26		

根据表 4 画出饼图如图，可以看出预期岗位中数据分析师和数据挖掘工程师这两类很热门。

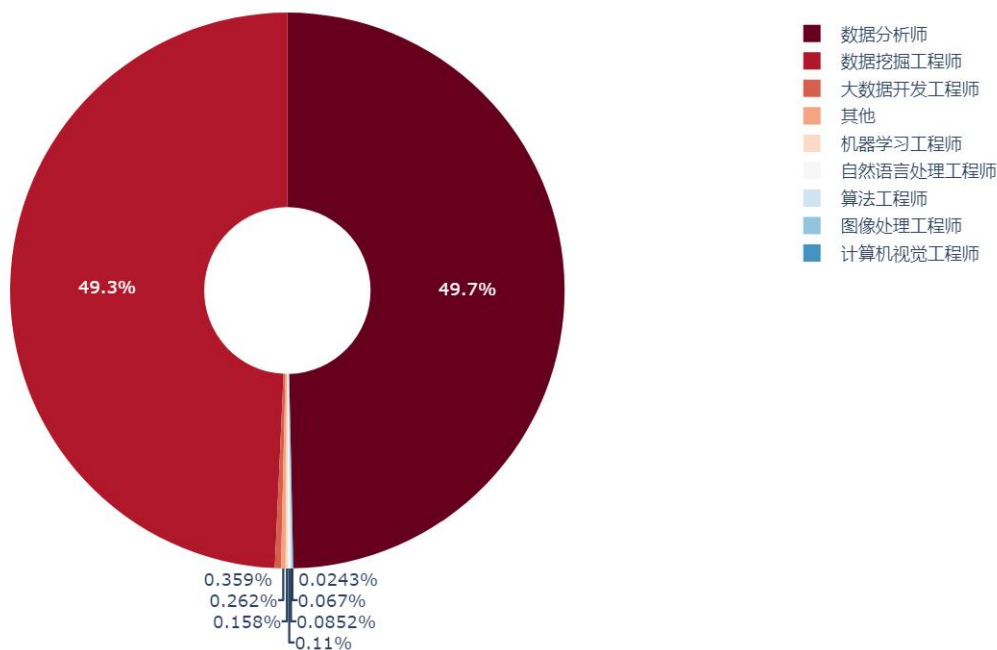


图 13 预期岗位饼图

2.期望行业

期望行业共有 13 种，绘制水平柱状图如图，可以看出互联网行业期望最大，同时公司行业数据中互联网行业也是需求最大。

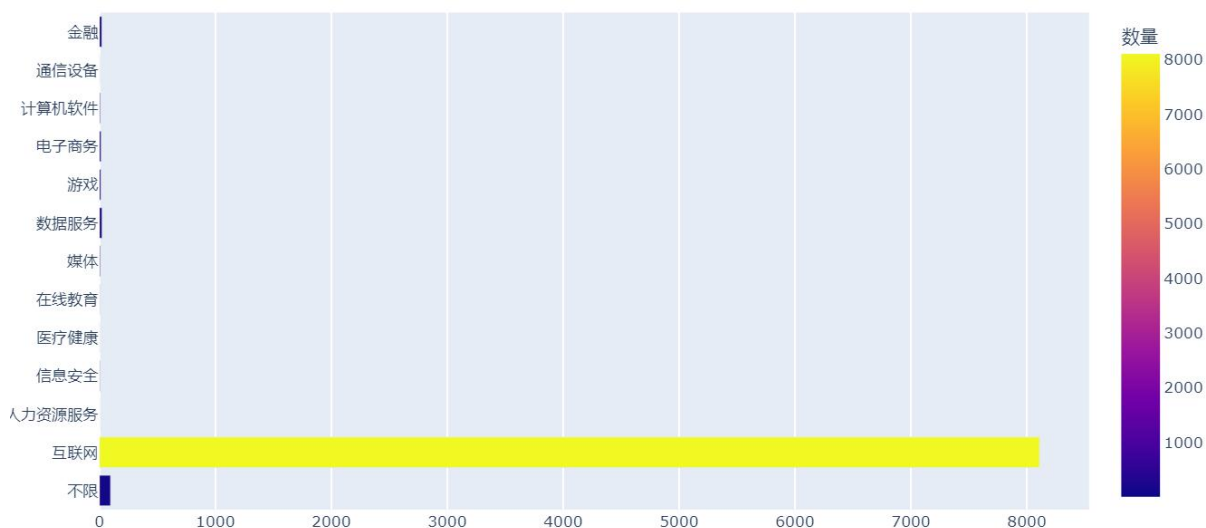


图 14 期望行业水平柱状图

3.薪资需求

求出预期最高薪资和预期最低薪资的平均值分别为 6069.63 和 4048.24，画出柱状图如图：

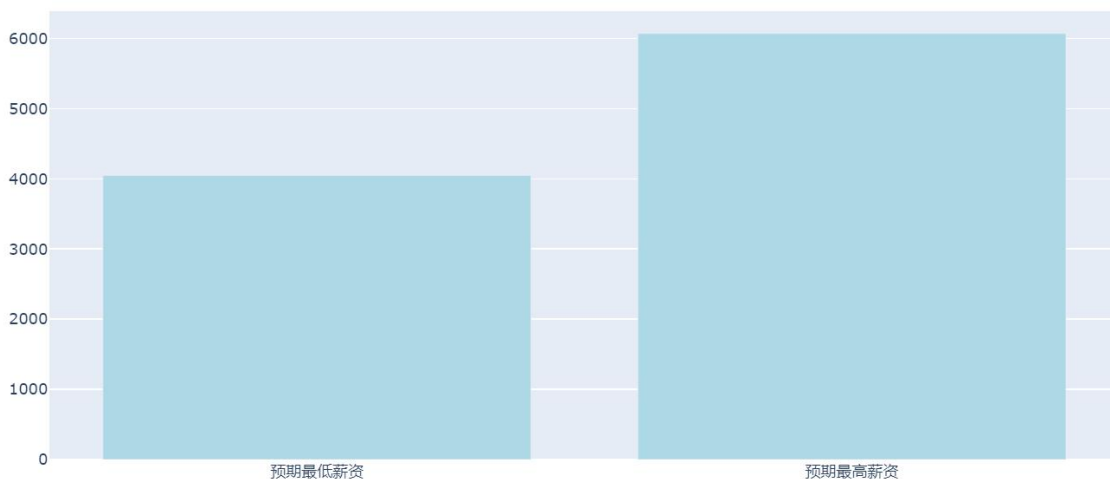


图 15 薪资需求柱状图

4.知识储备

使用简历关键词、技能、语言和证书数据来说明人才的知识储备。将各类数据进行分词，绘制词云图如下，可以看出知识储备比较多的有 Python、Pandas 库、Numpy 库、MySQL、Matplotlib 库、BeautifulSoup 库、scikit-learn 库。

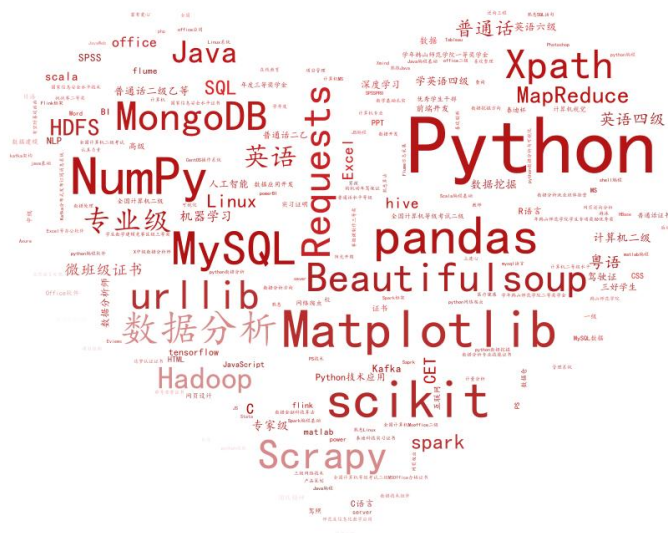


图 16 知识储备词云图

5. 学历和性别

根据学历数据和性别数据画出以下旭日图，可以看出男女比例接近 1:1，学历极大部分为本科。

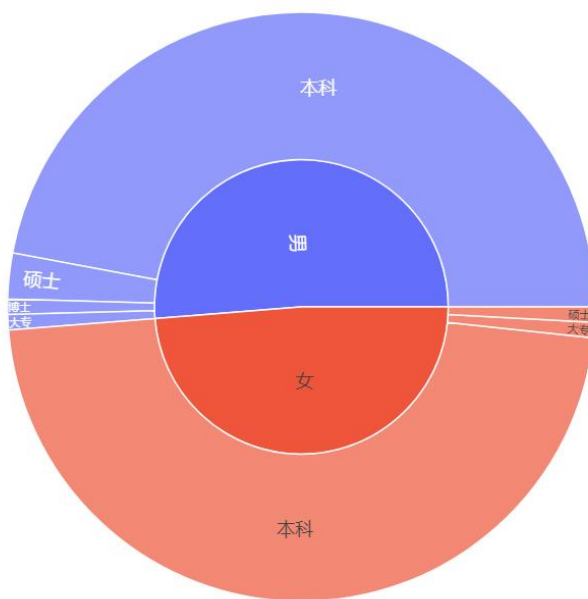


图 17 学历和性别旭日图

6. 工作经验

工作经验数据中只有 212 个有效数据，不具有代表性。

本节使用 Python 的 plotly 库对招聘信息和求职信息的多个方向进行了分析和可视化，得到了招聘信息画像和求职信息画像，并可以为求职者和企业提供更准确的参考信息。其中招聘信息画像总共从 8 个方向进行了分析，包括招聘岗位、学历要求、公司行业、公司类型、薪资待遇、企业工作地点、员工数量和岗位经验等方面。求职信息画像总共从 5 个方向进行了分析，包括预期岗位、期望行业、薪资需求、知识储备、学历和

性别等方面。

根据以上的招聘信息画像和求职信息画像，我们可以了解到市场需求有以下几点：

- (1) 数据分析师和数据挖掘工程师成为热门岗位，互联网行业是热门行业；
- (2) 广东省的招聘需求量更大，竞争也更大；
- (3) 企业对于岗位经验没有过多地要求，但更倾向于招聘岗位经验为 1-5 年的求职者。

我们可以为企业提供以下建议：

- (1) 在进行招聘时，可以考虑对学历层次不高、但技能关键字符合要求的求职者发出 offer；
- (2) 企业可以对不同岗位的招聘需求可以进一步细化，将会更加准确地招聘到对应的人才。

我们可以为求职者提供以下建议：

- (1) 大多数企业招聘要求学历为本科，所以求职者在学历方面，本科是最基本的要求，但如果想要提升自己的竞争力，求职者可以考虑继续深造并获得硕士和博士学位；
- (2) 求职者如果想要在岗位中有所突出，可以重点学习 Python 及相关库、数据分析等技能；
- (3) 求职者在进行简历投递时应该尽可能详尽地写出自己的相关信息，尤其是技能、工作经验和获奖情况等。

第三章 基于向量相似度的岗位匹配度和求职者满意度模型

3.1 模型假设

- (1) 假设企业和求职者的需求在招聘全过程中未发生改变；
- (2) 假设企业和求职者在招聘过程中完全理性，签约后不再改变选择；
- (3) 假设企业和求职者的所有条件都是硬性条件。

3.2 指标类型

由于人类掌握的测量技术等种种原因形成或制定的事物标准常常是多指标形式描述的。因此本文将描述标准的指标类型分为三种：01 判断信息型、区间值型和临界值型。[2][4]

(1) 01 判断信息型

01 判断信息，是指被评价主体中的个体在特定评价指标上的特征或属性是否与评价者要求的相一致，如果一致，则被评价者在这一指标上的得分为 1，否则为 0。

类似于员工“技能情况是否与企业技能要求相同”，岗位“是否提供食宿”，“是

否需要出差”等这类只能用“是”或“否”两种确定答案来回答的指标，均属于 01 判断信息型指标。[2]

(2) 区间值信息型

区间值信息型指标，是指被评价主体中的个体在某些特定指标上的回答，不能够直接用严格的是或否来评判，也不能准确的给定某一具体量化的数值，而只能是在某一范围内的取值。[2]

在实际的匹配决策中，所评价的指标涉及到的物理量纲不同，因此区间数所涉及的数值彼此差距很大，为消除不同物理量纲的影响，按照效益型和成本型将区间数进行规范化处理。效益型指标是指那些区间数值越大对评价者而言越有利的指标，而成本性指标是指区间数取值越小对评价者越有利的指标。[6]

诸如员工的“期望薪酬范围”，员工的“期望休假频率”等指标，通常都是无法用固定的具体实数来表示的，其取值常常是在一个区间内，这类指标均属于区间数信息型指标。

(3) 临界值信息型

临界值信息型指标，就是在确定的临界值下具有趋向性或越大越好或越小越好。由于越大越好和越小越好的临界值指标具有相互转换的特性以下只考虑越大越好的临界值指标。这类指标比较常见，如：某公司招聘秘书要求应聘人员英语水平在国家六级以上，那么这个六级以上的指标就属于临界值信息型指标。[5]

3.3 符号说明

本章中所涉及的主要符号及其含义如表 5 所示：

表 5 第三章主要符号对照表

符号名称	符号含义
A	企业招聘职位信息之集 $A=\{A_1, A_2, A_3, \dots, A_n\}$
B	求职者信息之集 $B=\{B_1, B_2, B_3, \dots, B_m\}$
$A_i, i=1,2,\dots,n$	第 i 个招聘职位的信息
$B_j, j=1,2,\dots,m$	第 j 个求职者的信息
R	企业招聘职位信息向量

I	求职者求职特征向量
$R_i=(R_{ik}, k=1,2,\dots,6)$	第 i 个招聘职位信息向量
$I_j=(I_{jk}, k=1,2,\dots,6)$	第 j 个求职者求职特征向量

注：表中只列举了文中部分符号，其它符号在文中给出。

3.4 岗位匹配度模型

3.4.1 招聘需求指标

每条招聘信息生成一个向量，该向量的各个分量为招聘企业的各个条件量化后的结果，所有招聘信息向量或求职意向向量各形成一个向量集，与招聘信息之集 A 和求职信息之集 B 一一对应。如，职位招聘信息向量：招聘(岗位，行业，学历，技能要求，工作经验，职位月薪，性别，年龄，工作地点，……)，个人求职信息向量：求职(预期岗位，预期行业，学历，技能关键字，工作经验，预期月薪，性别，年龄，期望工作地点，……)，企业招聘向量和求职者求职向量的各个分量也是一一对应的，如图 18 所示。生成向量时考虑到企业和求职者的实际情况及双向推荐所注重的因素，本模型向量只采用了岗位、行业、职位月薪、工作经验、技能要求、学历这 6 个分量。

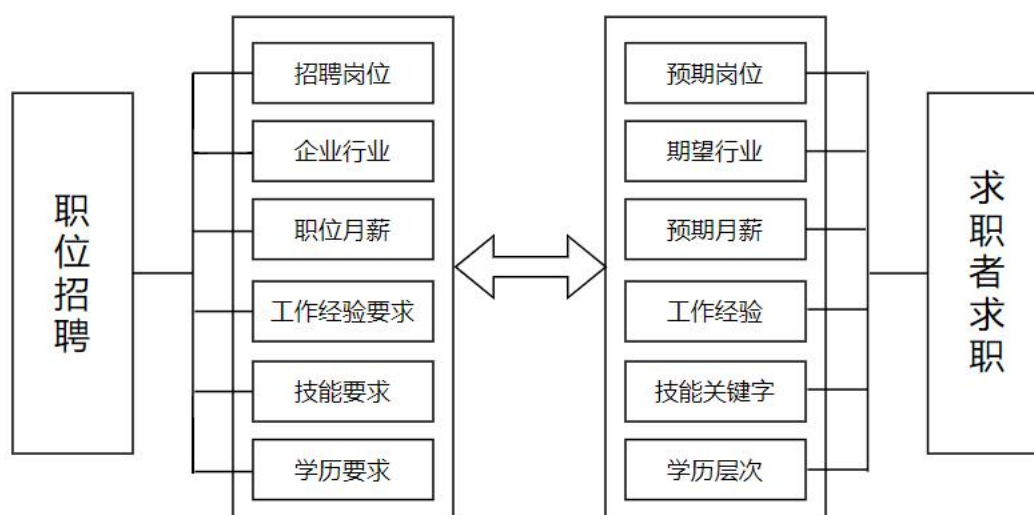


图 18 招聘求职信息指标对应图

企业招聘职位信息向量记为 $R=($ 招聘岗位 R_1 ，行业 R_2 ，职位月薪 R_3 ，工作经验 R_4 ，技能要求 R_5 ，学历 $R_6)$ 。求职者求职特征向量记为： $I=($ 预期岗位 I_1 ，预期行业 I_2 ，预期月薪 I_3 ，工作经验 I_4 ，技能关键字 I_5 ，学历 $I_6)$ 。

根据从泰迪内推平台上采集的职位信息与求职者信息，我们记职位信息全体 $A=\{A_1, A_2, A_3, \dots, A_n\}$ ，共有 n 条职位信息；记求职者信息全体 $B=\{B_1, B_2, B_3, \dots, B_m\}$ ，共有 m 条求职者信息。记第 j 个求职者对第 i 个职位的岗位匹配度为 S_{ij} , $i=1, 2, 3, \dots, n$, $j=1, 2, 3, \dots, m$ 。

3.4.2 招聘需求指标的量化方法

为每个职位推荐求职者时以职位的招聘信息向量为标准向量，其各个分量的值设为 1，对于个人求职信息向量，其每个向量的每个分量的取值范围为 $[0, 1]$ ，完全不匹配取值 0，完全匹配或者不限制取值为 1。对向量的各分量（建立映射或映射表）逐一进行量化，各分量的量化基本规则如下。

(1) 招聘岗位

当向企业推荐求职者时，在满足其他招聘条件的前提下，若求职者的预期岗位与该职位的招聘岗位相同则为 1，否则为 0。招聘职位要求的岗位为标准向量，其值为 1。

(2) 行业

当向企业推荐求职者时，在满足其他招聘条件的前提下，若求职者的预期行业与该职位行业相同则为 1，否则为 0。企业的行业为标准向量，其值为 1。

(3) 职位月薪

当向企业推荐求职者时，从企业招聘角度出发考虑，仅就薪酬这个因素而言，会优先考虑对薪酬期望值更低的求职者，因此在向企业推荐求职者时，将开出薪酬更低的求职者排在前面，因此构造以下量化规则。设企业薪酬最低值为 p ，求职者薪酬要求最低值为： p_1, p_2, \dots, p_n ，令 $M=\max\{p_1, p_2, \dots, p_n\}$, $m=\min\{p_1, p_2, \dots, p_n\}$ ，则企业薪酬为标准向量，其值为 1，求职者薪酬按公式 进行量化。

$$I_3 = \frac{p-x}{M-m} + \frac{M-p}{M-m}, x \in \{p_1, p_2, \dots, p_n\}$$

(4) 工作经验

当向企业推荐求职者时，就工作经验这个因素而言，如果企业没有特殊要求，将优先向企业推荐工作经验长的求职者；如果企业有要求，一般是最低限度的要求，将为企业推荐满足要求且工作经验更长的求职者，因此构造以下量化规则。

设企业要求工作经验最低值为 s 年，若求职者工作经验为 x 年，则

$$\mu(x) = \begin{cases} 1, & x \geq 10, \\ 0.1x, & 0 \leq x < 10. \end{cases}$$

构造简单函数

$$\varphi(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

则求职者工作经验按以下公式进行量化。

$$I_4 = \begin{cases} \mu(x), s = 0, \\ \varphi(x-s)\mu(x), s \geq 1. \end{cases}$$

其中，未填写工作经验的求职者看作工作经验为 0。

(5) 技能要求

当向求职者推荐岗位时，在满足其他招聘条件的前提下，若企业的职位关键字与求职者的求职关键字相同则为 1，否则为 0。职位的技能要求为标准向量，其值为 1。

(6) 学历要求

当向企业推荐求职者时，就学历这个因素而言，一般企业会有最低学历的要求，将为企业推荐满足要求且学历更高的求职者，根据此原则制定下表。

表 6 学历量化值对应表

学历	标准向量	匹配向量量化值
博士	1	1
硕士	1	0.8
本科	1	0.6
大专	1	0.4
技校	1	0.3
求职者未写学历	1	0.1
求职者无学历	1	0

设企业要求学历对应量化值最低值为 w ，若求职者学历对应量化值为 x ，则求职者学历按以下公式进行量化。

$$I_6 = \begin{cases} x, w = 0, \\ \varphi(x-w) \cdot x, w > 0. \end{cases}$$

3.4.3 相似度计算

计算相似度的方法有余弦相似度(Cosine Similarity)、皮尔逊相(Pearson)相似度和杰卡德(Jaccard)相似度等。[12]

(1) 余弦相似度(Cosine Similarity)

余弦相似度是将职位与求职者的匹配度或求职者对企业的满意度看成空间向量，要了解匹配度或满意度与求职者或企业的需求的相似程度，就可以通过计算它们与标准向量的余弦值，来衡量相似度大小，余弦值越大，两向量之间夹角越小，相似度越高。余弦相似度公式如下：[12]

$$sim(\mathbf{u}, \mathbf{v}) = \frac{|\mathbf{u} \cdot \mathbf{v}|}{|\mathbf{u}| |\mathbf{v}|}$$

(2) 杰卡德 (Jaccard) 相似度

杰卡德相似度是从集合角度去考虑，即二者相匹配的项目数占所有项目的比例。高相似度计算方法只注重“是”“否”，而忽略具体评分影响。一般用于离散型二元变量，对于非二元变量时的使用场景，计算效果较差。杰卡德相似度公式如下：[12]

$$sim_{i,j} = \frac{|u_{ij}|}{|u_i \cup u_j|}$$

(3) 皮尔逊 (Pearson) 相似度

皮尔逊相似度又称为相关相似度，是根据二者相匹配项目来计算线性相关程度的一种统计方法。皮尔逊相似度介于-1和1之间，当值为负数时，表示两向量负相关，为正数则呈正相关，值越接近上下限两向量线性关系越强。皮尔逊相似度公式如下：[12]

$$sim(u, v) = \frac{\sum_i (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_i (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_i (r_{v,i} - \bar{r}_v)^2}}$$

根据不同公式的适配度高低，本文采用余弦相似度的方法来计算相似度。

在为职位推荐求职者，需要计算他们之间的匹配度，匹配度根据3.4.2节的量化结果进行计算，即计算每条求职者推荐向量与标准向量之间的相似度，作为岗位匹配度，然后按这个岗位匹配度从高到低排序，使得更符合的求职者能排在前面。

向量相似度的计算采用夹角余弦公式来计算，计算公式如下所示。

$$\cos \langle R, I \rangle = \frac{R \cdot I}{|R| |I|}$$

其中，

$$\begin{cases} R \cdot I = R_1 I_1 + R_2 I_2 + \dots + R_6 I_6, \\ |R| \cdot |I| = \sqrt{R_1^2 + R_2^2 + \dots + R_6^2} \cdot \sqrt{I_1^2 + I_2^2 + \dots + I_6^2}. \end{cases}$$

在向职位推荐时， R 应该是职位信息标准向量。

3.4.4 基于向量相似度的岗位匹配度模型

(一) 模型思想

面向企业的岗位匹配度模型的总体思想是：首先从招聘企业的基本信息和招聘信息中提取有用的信息，构建企业招聘职位信息集合 $\{R_1, R_2, R_3, R_4, R_5, R_6\}$ ，得到每个职位的标准向量，计算每个求职者向量与招聘职位标准向量之间的相似度，以相似度作为岗位匹配度。具体步骤如下：

(1) 提取每个职位的招聘要求和企业特征。根据职位及其对应企业的基本信息和招聘需求，提取信息主要包括：招聘岗位，行业，职位月薪，工作经验，技能要求，学历，通过适当的量化方式将其转化为向量形式，以得到企业招聘职位信息集合 $\{R_1, R_2,$

$R_3, R_4, R_5, R_6\}$ ，得到每个职位的标准向量。

(2) 计算每个求职者与每个招聘职位标准向量之间的相似度，作为岗位匹配度。

(3) 生成排序列表。将步骤(2)中得到的岗位匹配度进行降序排序。

(二) 算法描述

根据上述算法思想，可得算法描述如下：

输入：招聘职位 A_i ，

输出：排序列表，

(1) 提取招聘职位 A_i 的属性特征，获得标准向量；

(2) 计算每个求职者在职位 A_i 要求下的特征向量；

(3) 计算每个求职者与职位招聘特征之间的相似度，作为岗位匹配度；

(4) 将 m 个求职者，按照匹配度大小进行排序，生成求职者排序列表。

3.5 求职者满意度模型

3.5.1 求职需求指标

根据从泰迪内推平台上采集的职位信息与求职者信息，我们记职位信息全体= $\{A_1, A_2, A_3, \dots, A_n\}$ ，共有 n 条职位信息；记求职者信息全体= $\{B_1, B_2, B_3, \dots, B_m\}$ ，共有 m 条求职者信息。记第 j 个求职者对第 i 个职位的满意度为 S_{ji} 。

3.5.2 求职需求指标的量化方法

对求职者而言，以求职者的求职信息向量为标准向量，其各个分量的值设为 1，对于筛选出来的企业招聘信息向量集，每个向量的每个分量的取值范围为 $[0, 1]$ ，完全不匹配取值 0，完全匹配或不限制取值为 1。对向量的各分量（建立映射或映射表）逐一进行量化，各分量的量化基本规则如下。

(1) 预期岗位

当向求职者推荐岗位时，在满足其他招聘条件的前提下，若职位的岗位与求职者的预期岗位相同则为 1，否则为 0。

(2) 预期行业

当向求职者推荐岗位时，在满足其他招聘条件的前提下，若职位的行业与求职者的预期行业相同则为 1，否则为 0。

(3) 预期月薪

当向求职者推荐岗位时，分析求职者对薪酬的心理期望，显然会更优先考虑开出薪酬更高企业，因此在向求职者推荐岗位时，将开出薪酬更高的企业排在前面，因此构造以下量化规则。

设求职者薪酬要求的最低值为 q ，企业薪酬最低值为： q_1, q_2, \dots, q_n ，令 $M=\max(q_1,$

$q_2, \dots, q_n)$, $m = \min(q_1, q_2, \dots, q_n)$, 则求职者薪酬为标准向量, 其值为 1, 职位薪酬按以下公式进行量化:

$$R_3 = \frac{y-q}{M-m} + \frac{q-m}{M-m}, y \in \{q_1, q_2, \dots, q_n\}.$$

(4) 工作经验

若向求职者推荐岗位时, 则在满足招聘条件的前提下, 优先向求职者推荐最接近求职者工作经验的企业, 比如, 求职者的工作经验为 5 年, 则筛选出来的企业对工作经验的要求为 5 年及以下的招聘信息, 在向求职者推荐时将工作要求为 5 年的排在前面, 其他工作经验依次往后排。

设 t 为求职者的工作经验年数, y 为职位要求的最低工作经验年数. 由于前文已构造量化函数

$$\mu(x) = \begin{cases} 1, & x \geq 10, \\ 0.1x, & 0 \leq x < 10, \end{cases}$$

和简单函数

$$\varphi(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

在量化时, 参照以上已构造的工作经验量化函数和简单函数, 则职位工作经验按以下公式进行量化。

$$\begin{cases} t \geq 1, R_4 = \begin{cases} \frac{\varphi(t-y)\mu(t)}{\mu(y)}, & \mu(y) > 0, \\ 1, & \mu(y) = 0. \end{cases} \\ t = 0, R_4 = \begin{cases} 0, & \mu(y) > 0, \\ 1, & \mu(y) = 0. \end{cases} \end{cases}$$

其中, 企业对工作经验要求为不限时记要求最低工作经验为 0。

(5) 技能关键字

当向求职者推荐岗位时, 在满足其他招聘条件的前提下, 若职位的职位关键字与求职者的求职关键字相同则为 1, 否则为 0。

(6) 学历层次

若向求职者推荐岗位, 则在满足招聘条件的前提下, 优先向求职者推荐最接近求职者学历的职位, 比如, 求职者的学历为本科, 则筛选出来的职位对学历要求为本科及以下的招聘信息, 在向求职者推荐时将学历要求为本科的排在前面, 其他学历依次往后排。

设 z 为求职者的学历层次量化值, y 为职位要求的最低学历量化值. 前文已构造简单函数

$$\varphi(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

和下述学历量化值对应表。

表 7 学历量化值对应表

学历	标准向量	匹配向量量化值
博士	1	1
硕士	1	0.8
本科	1	0.6
大专	1	0.4
技校	1	0.3
求职者未写学历	1	0.1
求职者无学历 或职位对学历要求不限	1	0

在量化时，参照以上已构造的简单函数和学历量化值对应表。若求职者有学历或已填写学历，优先推荐学历要求最接近的职位；若求职者未填写学历，优先推荐对学历要求不限或要求低的职位，学历要求越高的职位越不推荐。则职位学历要求按以下公式进行量化。

$$\left\{ \begin{array}{l} z > 0.1, R_6 = \begin{cases} \frac{z \cdot \varphi(z - y)}{y}, y > 0, \\ 1, y = 0. \end{cases} \\ z = 0.1, R_6 = \begin{cases} \frac{z}{y}, y > 0, \\ 1, y = 0. \end{cases} \\ z = 0, R_6 = \begin{cases} 0, y > 0, \\ 1, y = 0. \end{cases} \end{array} \right.$$

其中，职位对学历要求不限时记要求最低工作经验为 0。

3.5.3 相似度计算

在为职位推荐求职者，需要计算他们之间的匹配度，匹配度根据 3.5.2 节的量化结果进行计算，即计算每条职位推荐向量与标准向量之间的相似度，作为求职者，然后按这个求职者满意度从高到低排序，使得更符合的职位能排在前面。

向量相似度的计算采用夹角余弦公式来计算，计算公式如下所示。

$$\cos \langle R, I \rangle = \frac{R \cdot I}{|R||I|}$$

其中，

$$\begin{cases} R \cdot I = R_1 I_1 + R_2 I_2 + \dots + R_6 I_6, \\ |R| \cdot |I| = \sqrt{R_1^2 + R_2^2 + \dots + R_6^2} \cdot \sqrt{I_1^2 + I_2^2 + \dots + I_6^2}. \end{cases}$$

在向求职者推荐时， I 应该是求职者信息标准向量。

3.5.4 基于向量相似度的求职者满意度模型

(一) 模型思想

面向求职者的求职者满意度模型的总体思想是：首先从求职者的基本属性特征和求职需求中提取有用的信息，构建求职者的求职信息集合 $\{I_1, I_2, I_3, I_4, I_5, I_6\}$ ，以求职者的信息向量作为标准向量，计算招聘职位与求职者之间的相似度，以相似度作为求职者满意度。具体步骤如下：

(1) 提取每个求职者的招聘要求和企业特征。根据职位及其对应企业的基本信息和招聘需求，提取信息主要包括：预期岗位，预期行业，预期月薪，工作经验，技能关键字，学历层次，通过适当的量化方式将其转化为向量形式，以得到求职者的求职信息集合 $\{I_1, I_2, I_3, I_4, I_5, I_6\}$ 。

(2) 计算每个招聘职位与每个求职者之间的相似度，作为求职者满意度。

(3) 生成排序列表。将步骤 (2) 中得到的求职者满意度进行降序排序。

(二) 算法描述

根据上述算法思想，可得算法描述如下：

输入：求职者 B_j ，

输出：排序列表，

(1) 提取求职者 B_j 的求职特征，获得标准向量；

(2) 计算每个职位在求职者 B_j 需求下的特征向量；

(3) 计算每个职位与求职者求职特征之间的相似度，作为求职者满意度；

(4) 将 n 个职位，按照满意度大小进行排序，生成职位排序列表。

3.6 模型结果

根据前文的模型，我们可以求解出岗位匹配度和求职者满意度的结果，存放在 `resul3-1.csv` 和 `resul3-2.csv` 中，如图 19 和图 20：

	A	B	C
1	招聘信息ID	求职者ID	岗位匹配度
2	1648527394191052802	1468183620412899328	1
3	1648527394191052802	1468149322909614080	0.990672258
4	1648527394191052802	1645750326164324352	0.904172719
5	1648527394191052802	1461512488951611392	0.82135141
6	1648527394191052802	1639531141977473024	0.804331991

图 19 岗位匹配度结果示例

	A	B	C	D
1	求职者ID	招聘信息ID	公司名称	求职者满意度
2	1647944223779061760	1482195868793831427	云米全屋互联网	1
3	1647944223779061760	1482185135670558721	网易集团	1
4	1647944223779061760	1482195642796343302	云米全屋互联网	1
5	1647944223779061760	1482184690793316354	有米科技	1
6	1647944223779061760	1482192783140847622	广州阳光喔教育	0.993553023
7	1647944223779061760	1482196338836897796	Flow++	0.980576601

图 20 求职者满意度结果示例

第四章 招聘求职双边匹配推荐系统

4.1 算法选择

根据招聘流程，我们发现，向企业推荐求职者时需要充分考虑岗位匹配度、求职者满意度和履约率三者。

(1) 为了使履约率尽可能高，需要对推荐的方式尽可能优化。更进一步，应该充分考虑求职者和岗位的选择策略，从而达到优化的目标。同时，招聘流程可以简单看做是一轮又一轮的选择，若在每一轮中，系统都能为企业推荐最适合的求职者，最终招聘流程结束时，就会得到整体最适合的人岗匹配情况。因此，我们考虑采用贪心算法建立推荐模型，将每一轮招聘都看做是一个子问题，求解出每一个子问题的最优解，从而得到整体最优解。

(2) 由于贪心算法一定程度上十分依赖贪心策略的选择，如果可以建立一个更为准确而理性的模型来获取最终的人岗匹配推荐情况，将会使系统的推荐精确性和效率更高。而且，一味地考虑履约率并不符合实际的企业招聘和求职者求职情况，最理想的状态应该是企业能招到与职位匹配度尽可能高的求职者，而求职者能被自己满意的职位录用。在考虑这两方面的情况下，尽可能提高履约率，才是一个运行良好的推荐系统。

因此，我们考虑建立多目标规划模型，通过模型求解，找出使岗位匹配度、求职者满意度和履约率达到最高的人岗匹配情况，该情况就是系统应该对企业推荐的求职者情况。

4.2 符号说明

本章中所涉及的主要符号及其含义如表 8 所示：

表 8 第四章主要符号对照表

符号名称	符号含义
C_j	求职者 B_j 最终选择的岗位
S_j	求职者 B_j 收到的达到“合格值”的 offer
l_i	岗位 A_i 拟招聘人数
x_{ij}	求职者 B_j 是否被职位 A_i 录用

注：表中只列举了本章中常用的部分符号，其它符号在文中给出。

4.3 算法理论基础

4.3.1 贪心算法

贪心算法，又称贪婪算法，是指在对问题求解时，总是做出在当前看来是最好的选择。也就是说，不从整体最优上加以考虑，算法得到的是在某种意义上的局部最优解。贪心算法不是对所有问题都能得到整体最优解，关键是贪心策略的选择。[7]

贪心算法一般按如下步骤进行：

- ①建立数学模型来描述问题。
- ②把求解的问题分成若干个子问题。
- ③对每个子问题求解，得到子问题的局部最优解。
- ④把子问题的解局部最优解合成原来解问题的一个解。

贪心算法是一种对某些求最优解问题的更简单、更迅速的设计技术。贪心算法的特点是一步一步地进行，常以当前情况为基础根据某个优化测度作最优选择，而不考虑各种可能的整体情况，省去了为找最优解要穷尽所有可能而必须耗费的大量时间。

贪心算法采用自顶向下，以迭代的方法做出相继的贪心选择，每做一次贪心选择，就将所求问题简化为一个规模更小的子问题，通过每一步贪心选择，可得到问题的一个最优解。虽然每一步上都要保证能获得局部最优解，但由此产生的全局解有时不一定是最优的，所以贪心算法不要回溯。[8]

4.3.2 多目标规划

多目标规划是数学规划的一个分支，研究多于一个目标函数在给定区域上的最优化。又称多目标最优化。通常记为 VMP。在很多实际问题中，例如经济、管理、军事、科学和工程设计等领域，衡量一个方案的好坏往往难以用一个指标来判断，而需要用多个目标来比较，而这些目标有时不甚协调，甚至是矛盾的。[9]

求解多目标规划常用方法是把多目标规划问题转化为单目标的数学规划问题进行求解，即标量化。[10]

任何多目标规划问题，都由两个基本部分组成：（1）两个及以上的目标函数；（2）若干个约束条件。有 n 个决策变量， k 个目标函数， m 个约束方程。则 $Z=F(X)$ 是 k 维函数向量， $\Phi(X)$ 是 m 维函数向量； G 是 m 维常数向量。[10]

4.3.3 整数规划

整数规划是指规划中的变量（全部或部分）限制为整数，若在线性模型中，变量限制为整数，则称为整数线性规划。所流行的求解整数规划的方法往往只适用于整数线性规划。从约束条件的构成，整数规划又可细分为线性、二次和非线性的整数规划。[11]

在线性规划问题中，有些最优解可能是分数或小数，但对于某些具体问题，常要求某些变量的解必须是整数。例如，当变量代表的是机器的台数，工作的人数或装货的车数等。为了满足整数的要求，初看起来似乎只要把已得的非整数解舍入化整就可以了。实际上化整后的数不见得是可行解和最优解，所以应该有特殊的方法来求解整数规划。在整数规划中，如果所有变量都限制为整数，则称为纯整数规划；如果仅一部分变量限制为整数，则称为混合整数规划。整数规划的一种特殊情形是 01 规划，它的变数仅限于 0 或 1。[11]

4.4 基于贪心算法与策略优化的招聘求职双边匹配模型

在招聘的过程中，条件优秀的岗位或求职者都受到青睐，根据目前的招聘流程，如果企业没有按照合适的策略向求职者发出 offer，最终招聘的情况可能会不太理想，导致招聘人数不足或招聘的求职者与岗位匹配度不高等问题。因此，系统对企业推荐求职者的情况就至关重要了。对应的，评价推荐系统优劣的一个重要指标就是履约率。

履约率=所有岗位的签约人数之和/所有拟聘岗位人数之和

由于所有拟聘岗位人数之和是相对固定的，求职者数目也是固定的，为了使得履约率尽可能高，只需确保每个岗位能尽量招满人即可。而在招聘流程中，与之相关的最重要的问题就是系统应按照何种策略向企业推荐适合各岗位且签约概率高的求职者，因此我们对推荐策略进行分析与优化，使得履约率最高。

4.4.1 求职者选择策略

在招聘流程中，当求职者收到企业的 offer，将选取满意度最高的岗位进行签约。需要考虑到有以下两个方面：

（1）在这些 offer 中，可能存在满意度很低的情况，因此需要对岗位进行筛选，设定一个满意度临界值，记为“合格值”，只有满意度超过合格值的岗位，求职者才会接受其 offer 并进行签约。

(2) 在这部分 offer 中, 可能会有求职者满意度相同的情况, 此时, 求职者优先选取薪酬更高的岗位。

综合考虑, 针对求职者选择环节, 基于贪心算法可构造一个选择函数, C_j 表示求职者 B_j 最终选择的岗位, S_j 表示求职者 B_j 收到的达到“合格值”的 offer, 则有

$$C_j = \arg \max_{i \in S_j} \beta_{ij}.$$

4.4.2 岗位选择策略

(1) **策略一:** 为了招聘符合企业要求的员工, 企业更倾向于招聘匹配度高的求职者, 因此为了达到此目标, 将根据匹配度结果向职位 A_i 推荐匹配度前 l_i 的求职者, 企业将对这些求职者发出 offer。

(2) **策略二:** 在策略一的基础之上, 在 offer 发出之前, 实际上系统已经得出每位求职者与每个岗位的岗位匹配度与求职者满意度, 如果存在某些职位与求职者互为对方的匹配度与满意度最高, 那么这部分职位与求职者可以直接进行匹配, 即系统应优先考虑向企业推荐满足这样条件的求职者。而对于其他职位, 系统在进行推荐时则应当剔除这部分求职者。因此, 一方面, 系统应先筛选出匹配度和满意度互为最高的求职者与企业, 进行匹配推荐; 另一方面, 若同一批求职者与其他职位的岗位匹配度也是最高, 则应在对其他职位进行推荐时剔除同一批求职者, 以保证 offer 的可用性与有效性。

因此, 策略二的推荐步骤应为:

Step1. 筛选出互为最高的求职者与职位, 向职位进行推荐;

Step2. 在进行推荐时, 对于出现“双向奔赴”的情况的职位, 按照策略一进行推荐; 对于未出现“双向奔赴”情况的职位, 剔除 Step1 中的求职者, 再按照岗位匹配度从高到低排序, 向职位 A_i 推荐匹配度前 l_i 的求职者。

(3) **策略三:** 在策略一和策略二中, 我们将职位的岗位匹配度视为最重要的因素, 优先考虑了求职者是否满足职位的要求, 但实际上, 求职者的满意度同样需要进行考虑。在招聘过程中, 实际上一定会出现企业希望招聘的求职者有更满意的其他选择, 在此情况下企业很大可能会被拒绝, 在进行第二轮招聘时, 可选择的求职者将大大减少, 最终招聘到的求职者与职位的匹配度可能会比较低, 且也可能导致剩余的求职者对职位的满意度不高, 这样签约人数将会减少, 使得履约率降低。

因此为了提高履约率, 我们考虑优先考虑求职者的满意度, 先选出对职位 A_i 满意度最高的求职者, 考虑在其中选出与该职位岗位匹配度最高的前 l_i 位求职者进行推荐。

在此策略下, 考虑到以下两个问题: 若对职位满意度最高的求职者人数未超过职位招聘人数, 应该如何解决; 若对职位满意度最高的求职者与此职位的匹配度太低, 则会出现职位招聘质量过低的情况。所以, 我们也为职位设定匹配度“合格值”, 只有求职者与职位的匹配度超过“合格值”才可以被系统推荐至企业。

因此, 策略三的推荐步骤应为:

Step1. 选出所有对职位 A_i 满意度最高的求职者；若求职者人数比职位招聘人数多，则进行 Step2；若求职者人数比职位招聘人数少，则将满意度排名加 1，选出所有对职位 A_i 满意度前 2 的求职者，以此类推，直至数量满足要求。

Step2. 按照求职者与该职位的岗位匹配度从高到低排序，剔除未达“合格值”的求职者；若剔除后求职者人数比职位招聘人数多，则继续进行 Step3；若剔除后求职者人数比职位招聘人数少，则回到 Step1 将满意度降 1 进行筛选，依次类推，直至数量满足要求；

Step3. 在达到“合格值”的求职者中，向职位 A_i 推荐匹配度前 l_i 的求职者。

4.4.3 招聘求职双边匹配推荐模型

在招聘流程中，若某一职位在第一轮未招满员工，将会继续进行下一轮招聘，直至招满为止。

根据贪心算法的思想，本文将每一轮招聘都看做一个子问题，综合求职者选择策略和岗位选择策略，制定三种策略下的招聘步骤，使得每一轮招聘都能达到最优的情况，从而找到全局最优的情况。不同策略下的算法步骤如下：

策略一：

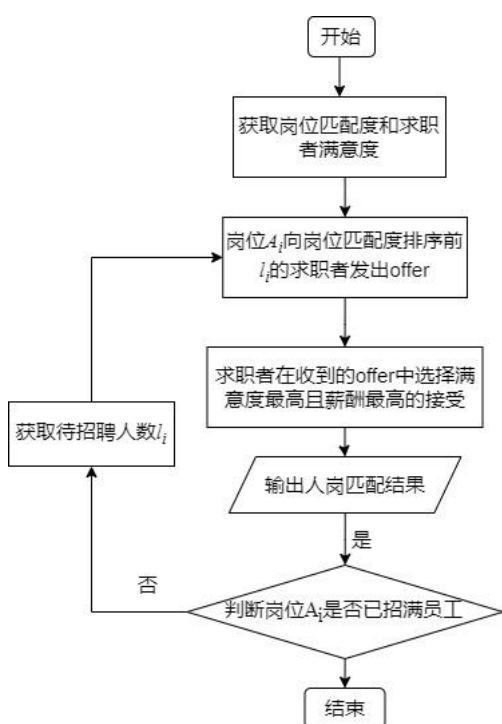


图 21 策略一算法流程图

策略二：

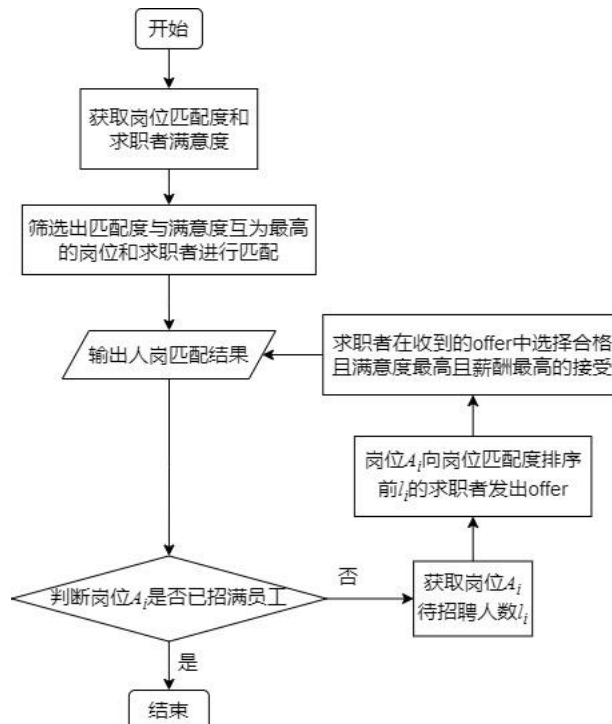


图 22 策略二算法流程图

策略三:

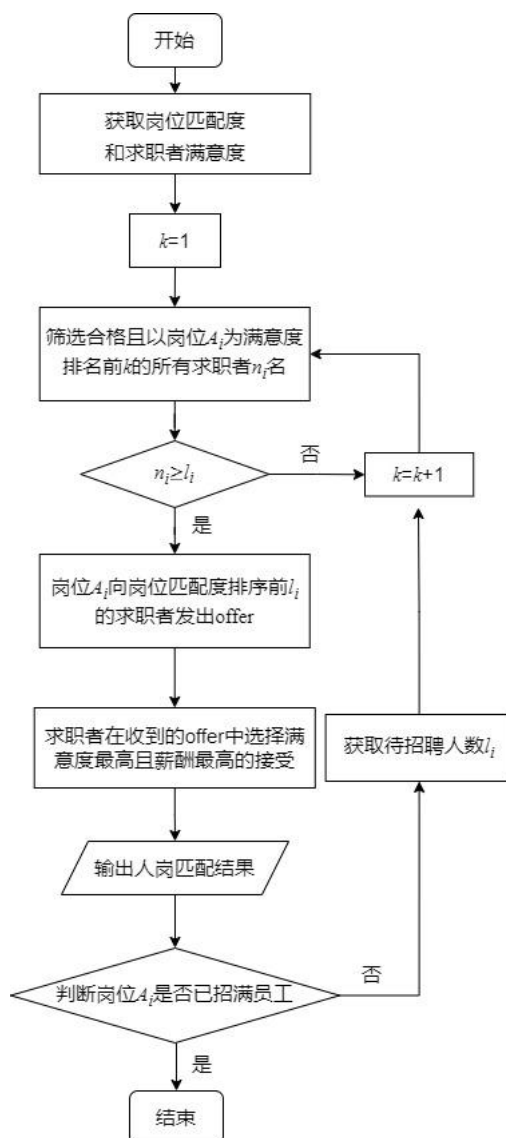


图 23 策略三算法流程图

4.5 基于多目标规划的招聘求职双边匹配推荐模型

在招聘的过程中，条件优秀的岗位或求职者都受到青睐，根据目前的招聘流程，如果企业没有按照合适的策略向求职者发出 offer，最终招聘的情况可能会不太理想，导致招聘人数不足或招聘的求职者与岗位匹配度不高等问题。因此，系统对企业推荐求职者的情况就至关重要了。对应的，评价推荐系统优劣的一个重要指标就是履约率。

履约率=所有岗位的签约人数之和/所有拟聘岗位人数之和

由于所有拟聘岗位人数之和是相对固定的，求职者数目也是固定的，为了使得履约率尽可能高，只需确保所有岗位的签约人数尽可能多，也就是成功签约的求职者人数尽可能多。

鉴于此，本章在充分考虑到人岗双方的匹配度和满意度的前提下，建立人岗双方的

感知效用矩阵，并以感知效用为依据，通过多目标规划模型的建立和求解，得到一套使签约的求职者人数与人岗双方感知效用之和最大的匹配决策方案。

4.5.1 感知效用矩阵构建

要构建职位和求职者双方的感知效用矩阵，首先要得到岗位匹配度和求职者满意度，可应用第三章中的模型得出结果。记 α_{ij} 为职位 A_i 与求职者 B_j 的匹配度， β_{ij} 为求职者 B_j 对职位 A_i 的满意度。

由于在本模型中，最重要的是考虑岗位匹配度和求职者满意度，因此本模型以 $(\alpha_{ij})_{n \times m}$ 和 $(\beta_{ij})_{n \times m}$ 作为双方的感知效用矩阵。

4.5.2 决策模型构建

为了使系统推荐的人岗匹配情况可以使履约率达到最高，我们要使得所有岗位的签约人数之和达到最大。

根据职位对求职者的感知效用矩阵 $(\alpha_{ij})_{n \times m}$ 和求职者对职位的感知效用矩阵 $(\beta_{ij})_{n \times m}$ ，可以建立一个使双方感知效用最大的多目标规划模型。设 x_{ij} 为0-1变量，表示求职者 B_j 是否被职位 A_i 录用。当 $x_{ij}=1$ 时，表示求职者 B_j 被职位 A_i 录用；当 $x_{ij}=0$ 时，表示求职者 B_j 未被职位 A_i 录用。决策模型如下：

$$\begin{aligned} \max Z_A &= \sum_{i=1}^n \sum_{j=1}^m \alpha_{ij} x_{ij} \\ \max Z_B &= \sum_{i=1}^n \sum_{j=1}^m \beta_{ij} x_{ij} \\ \max Z_C &= \sum_{i=1}^n \sum_{j=1}^m x_{ij} \\ \text{s.t.} &\begin{cases} 0 \leq \sum_{j=1}^m x_{ij} \leq l_i, 1 \leq j \leq m \\ 0 \leq \sum_{i=1}^n x_{ij} \leq 1, 1 \leq i \leq n \\ x_{ji} \in \{0, 1\}, 1 \leq j \leq m, 1 \leq i \leq n \end{cases} \end{aligned}$$

其中，决策变量 x_{ij} 表示第 j 个求职者是否被第 i 个职位录用，若被录用，则 $x_{ij}=1$ ；否则 $x_{ij}=0$ 。

三个约束条件中：

第一条保证第 i 个职位招聘的人数不超过拟聘人数；

第二条保证第 j 个求职者最多只能录用到一个职位；

第三条保证第 j 个求职者最多只能被一个职位录用一次。

4.5.3 决策模型求解

对于构建的模型，我们发现三个目标的系数矩阵取值范围都在 $[0, 1]$ 之间，即三者属于同一量纲。为方便计算，对三个目标进行线性加权，建立单目标规划模型，并对其进行求解。

$$\begin{aligned} \max Z &= \omega_A \sum_{i=1}^n \sum_{j=1}^m \alpha_{ij} x_{ij} + \omega_B \sum_{i=1}^n \sum_{j=1}^m \beta_{ij} x_{ij} + \omega_C \sum_{i=1}^n \sum_{j=1}^m x_{ij} \\ \text{s.t.} &\begin{cases} 0 \leq \sum_{j=1}^m x_{ij} \leq l_i, 1 \leq j \leq m \\ 0 \leq \sum_{i=1}^n x_{ij} \leq 1, 1 \leq i \leq n \\ x_{ij} \in \{0, 1\}, 1 \leq j \leq m, 1 \leq i \leq n \end{cases} \end{aligned}$$

其中， ω_A 、 ω_B 和 ω_C 表示岗位匹配度、求职者满意度和签约人数在线性加权中的权重。在进行招聘求职双边匹配推荐的过程中，更看重哪一方面，即哪一方面掌握主导权，在其中所占的权重就比较大。且 ω_A 、 ω_B 和 ω_C 应满足：

$$0 \leq \omega_A \leq 1, 0 \leq \omega_B \leq 1, 0 \leq \omega_C \leq 1, \omega_A + \omega_B + \omega_C = 1.$$

在模型中，更看重签约人数尽可能多，因此 ω_C 最大；而为公平起见，岗位匹配度和求职者满意度应同等重要，因此 $\omega_A = \omega_B$ ；因此有 $\omega_C > \omega_A = \omega_B$ 。分别考虑以下几种情况的模型进行求解：

$$\begin{aligned} \omega_C &= 0.5, \omega_A = \omega_B = 0.25 \\ \omega_C &= 0.6, \omega_A = \omega_B = 0.2 \\ \omega_C &= 0.7, \omega_A = \omega_B = 0.15 \end{aligned}$$

由于目标函数和约束条件均为线性，故采用线性规划方法对模型进行求解，还可利用LINGO, MATLAB等软件求解该线性规划模型。

4.5.4 决策步骤整理

招聘求职人岗双边匹配决策步骤如下：

- (1) 得到岗位匹配度和求职者满意度；
- (2) 求出双方的感知效用矩阵，分别是岗位匹配度矩阵和求职者满意度矩阵；
- (3) 依据双方的感知效用矩阵，建立感知效用最大化和签约人数最大化的多目标规划模型；
- (4) 利用线性加权，将多目标规划模型转化为单目标规划模型，通过模型求解，获取匹配推荐方案。

4.6 模型结果

对于岗位需求量为不限的招聘信息，我们采取以下的处理方法：

$$A = (\text{求职者总人数} - B) / m$$

其中 A、m 分别为为岗位需求量为不限的招聘信息所需招聘量、招聘信息数量，B 为岗位需求量为具体数字的所有招聘信息所需招聘量。

所以岗位需求量为不限的招聘信息所需招聘量为 $(8229 - 5220) / 120 = 22.575$ ，向上取整得到 23，所以岗位需求量为不限的招聘信息所需招聘量为 23 个。

根据前文两种模型，我们可以求解出最终履约率最高的结果，存放在 resul4.csv 中，如图所示：

	A	B	C	D
1	招聘信息ID	求职者ID	岗位匹配度	求职者满意度
2	1570688591800172544	1645385086935367680	0.652065733	0.977318918
3	1493425099607506944	1639549004641599488	0.630530561	0.630992041
4	1648527394191052800	1627308899029876736	0.231166099	0.862432572

图 招聘求职双边匹配推荐模型结果示例

结果显示总计 8280 个招聘需求量，4631 个招聘需求量被满足，总计 8229 个求职者，4631 个求职者找到工作，总计 1575 个招聘信息，1252 个招聘岗位已招满。计算履约率为 $\text{履约率} = 4631 / 8229 \approx 0.5628$ 。

第五章 总结与展望

5.1 总结

本文通过爬虫获取招聘信息和求职信息数据，从多个方向建立招聘信息画像和求职者画像，为求职者提供一些建议。进而构建匹配度与求职者满意度模型，并基于该模型，为每条招聘信息提供岗位匹配度非 0 的求职者，以及为每位求职者提供求职者满意度非 0 的招聘信息。最后，分别基于贪心算法和多目标规划模型建立招聘求职双向匹配推荐模型，使得履约率达到 56.28%。

5.2 优缺点分析

5.2.1 优势

(1) 本文匹配度与求职者满意度模型考虑 6 个重要的指标，且不同指标的重要性不同，比较全面考虑了招聘信息和求职者的需求和实际情况；

(2) 本文招聘和求职双向推荐模型综合考虑了招聘信息的匹配度和求职者的满意

度，对招聘信息和求职者进行双向推荐，相比传统的单向推荐，本模型的双向推荐算法更具有实用性和全面性；

(3) 本文对招聘求职双向推荐问题提出了多种算法，从中选取了最优的招聘和求职双向推荐模型，说明该模型具有良好效果。

5.2.2 不足

(1) 由于求职者信息数据缺失值较多，所以部分指标没有使用，如：工作地区、年龄、到岗时间等。后续模型优化可以考虑更多的指标，为企业和求职者提供更合适的人选，提高模型履约率；

(2) 在构建岗位匹配度和求职者满意度模型时，未考虑进行聚类分析，可能导致相似度计算的准确度有所下降。

5.3 展望与优化

(1) 本文使用的数据为某日爬取的数据，而实际上网页数据是在不断更新的，可以将双向推荐模型推广成一个动态模型来适应动态数据。

(2) 该模型可以进行招聘和求职的双向推荐，也可以推广到其他领域，如客户与商品双向推荐、交友匹配等。

(3) 在构建岗位匹配度和求职者满意度模型时，可以进行聚类分析，比较结果是否有所优化。

(4) 应用贪心算法建立招聘求职双向匹配模型时，对岗位和求职者的选择策略加以改进，更加全面地进行考虑，将会得到更好的人岗匹配推荐方案。

参考文献

- [1]刘艳.基于协同过滤算法实现高校个性化就业推荐系统研究[J].现代信息科技,2019,3(15):10-11+14.
- [2]朱丽娜.人岗双边匹配决策方法研究[D].河北大学,2015.
- [3]揭正梅.基于协同过滤的高校个性化就业推荐系统研究[D].昆明理工大学,2015.
- [4]刘兴林,吴明芬,刘利伟.基于向量相似度的招聘就业双向推荐模型[J].中国科技信息,2013(21):174-179.
- [5]陈希,樊治平.组织中员工与岗位匹配的两阶段测评与选择方法[J].东北大学学报(自然科学版),2009,30(09):1337-1340.
- [6]赵希男,温馨,贾建锋.组织中人岗匹配的测算模型及应用[J].工业工程与管理,2008(02):112-117.DOI:10.19495/j.cnki.1007-5429.2008.02.025.
- [7]康桂花.计算概论[M].中国铁道出版社,2016.08,123.
- [8]谢翌,江渝川.大学计算机计算思维与应用[M].重庆大学出版社,2017.01,58.
- [9]李光金,刘永清.基于多目标规划的DEA[J].系统工程理论与实践,1997,17(3):17-23.
- [10]胡毓达.多目标规划有效性理论[M].上海科学技术出版社,1994.
- [11]黄力伟,冯杰,王勤,尹成义.军事运筹学[M].国防工业出版社,2016.01.
- [12]王晨璐.求职招聘双向推荐系统的设计及实现[D].河北农业大学,2017.