

多算法融合的就业推荐算法研究

黄虹霞, 曾军, 曾历

江西科技学院信息工程学院, 江西 南昌 330000

摘要: 本研究聚焦于多算法融合的就业推荐算法。综合运用基于用户、岗位的协同过滤融合算法, 并结合 TF-IDF 构建用户-岗位特征矩阵能够更好地捕捉用户和岗位的文本特征。此外, 引入 AutoRec 模型对用户-岗位特征数据进行低维特征提取与重构。经实验验证, 融合协同过滤推荐和 AutoRec 模型的算法在推荐效果上更优于单一的协同过滤推荐算法, 能够为实际就业推荐系统提供了更优的解决方案。

关键词: 算法融合; TF-IDF; 就业推荐; AutoRec 模型

Research on Employment Recommendation Algorithm with Multi Algorithm Fusion

Huang Hongxia, Zeng Jun, Zeng Li

Jiangxi University of Technology, Nanchang, Jiangxi 330000

Abstract: This study focuses on the employment recommendation algorithm that integrates multiple algorithms. The comprehensive application of collaborative filtering and fusion algorithms based on users and positions, combined with TF-IDF to construct a user position feature matrix, can better capture the textual features of users and positions. In addition, the Autorec model is introduced to extract and reconstruct low dimensional features from user job feature data. Through experimental verification, the algorithm that integrates collaborative filtering recommendation and Autorec model has better recommendation performance than a single collaborative filtering recommendation algorithm, and can provide a better solution for practical employment recommendation systems.

Keywords: algorithm fusion; TF-IDF; employment recommendations; AutoRec model

引言

面对瞬息万变的就业市场, 毕业生就业难题与企业招聘效率的低下已成为社会热点议题。据教育部统计, 2022至2023年间, 我国高校毕业生人数激增7.62%, 达1158万, 而就业岗位增长率仅为0.35%, 致使岗位竞争异常激烈。企业在招聘过程中同样面临严峻挑战。同时, 企业在招聘过程中也面临着效率低下、成本上升和新员工流失率高等问题。

这一背景下个性化就业推荐系统能够根据求职者的学历背景、实习经历、技能专长以及求职网站的行为偏好等信息, 精准推送适配岗位, 同时帮助企业通过解析岗位需求和企业文化特点, 筛选出高契合度的候选人, 从而显著提升双方的匹配效率。

国内外学者对就业推荐系统的研究不断深入。当前研究主要集中在利用大数据、人工智能和机器学习技术来提升推荐系统的准确性和效率。国内研究者如薛亮^[1]等人, 通过标签聚类和协同过滤算法, 探索了如何组织用户和职位数据以提高推荐准确性。韩义畴^[2]等人则分析了高职院校学生就业能力现状, 并提出了相应的培育对策。蒋大锐^[3]等人提出了基于模糊层次聚类的大学生就业数据分类存储系统, 旨在提高数据检索和分析的效率。李龙^[4]等人研究了基于改进 TF-IDF 算法的毕业生就业推荐算法, 而崔秀艳^[5]探讨了结合大数据分析和人工智能的智慧就业推荐系统设计与实现。在深度学习领域, 江宜橙^[6]研究了基于深度学习的高校学生就业推荐系统, 而金铄^[7]则专注于基于用户画像的高校毕业生就业推荐方法。国际上, Wan P.^[8]提出了基于 K-Means 改进的协同过滤算法的就业推荐系统, 通过聚类相似用户和改进相似度计算来预测求职者评分, 提高了推荐系统的精确度、召回率和 F-score 比率。Huang S^[9]等人则提出了一种用于制造业推荐系统的长命令子序列算法, 利用相似连接技术优化推荐系统性能。

综上所述, 国内外的研究者都在探索如何通过不同的算法和技术来提升就业推荐系统的性能。本文将在此基础上进一步研究多算法融合的就业推荐算法, 以提高推荐系统的准确性和用户满意度, 推动就业推荐系统向更高效、精准的方向发展。

基金项目: 江西科技学院自然科学技术项目, 课题题目: 多算法融合的就业推荐算法研究 (项目编号 24ZRZD06)。

作者简介:

黄虹霞 (1994-), 女, 江西省南昌市, 江西科技学院信息工程学院, 硕士研究生, 主要研究机器学习与数据挖掘等方向。

曾军 (1996-), 男, 江西省吉安人, 江西科技学院信息工程学院, 硕士研究生, 主要研究图像处理、深度学习等方向。

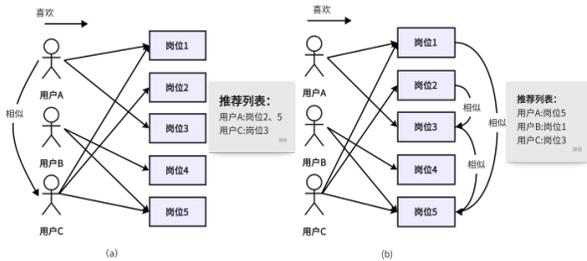
曾历 (1994-), 女, 江西省丰城人, 江西科技学院信息工程学院, 硕士研究生, 主要研究网络管理与优化、软件工程等方向。

一、推荐系统研究与算法概述

就业推荐系统旨在为求职者精准匹配职位，同时助力企业发掘合适的候选人。推荐系统依托用户历史行为、个人资料、简历等数据，运用算法预测用户潜在职位兴趣。主要算法包括协同过滤、TF-IDF及 AutoRec 模型推荐。

(一) 协同过滤推荐算法

协同过滤^[10,11]是一种重要的推荐技术，主要侧重于用户或岗位间的相似性分析。如图1所示，基于用户（U-CF）(a)或岗位（I-CF）(b)的协同过滤算法通过分析用户和岗位之间的相似性，为用户精准推荐岗位。



>图1 协同过滤推荐示意图

基于用户的协同过滤通过分析用户间的相似性，预测用户可能对职位的兴趣。相似性分析使用余弦相似度计算，公式如下：

$$\text{sim}(u, v) = \cos(\theta) = \frac{V_u \cdot V_v}{|V_u| \cdot |V_v|} \quad (1)$$

其中， u 和 v 是用户， V_u 和 V_v 为用户的特征和行为数据，包括学历、专业、工作经验、技能和对岗位的浏览、收藏、点赞等行为。上式计算结果大小意味着两用户之间的相似程度高低，因此，对用户集合通过相似度公式计算后得到的多个结果进行降序排序，并选取一定数量的用户，便可以组成相似用户集合 m^* 。

同理，基于岗位的协同过滤通过分析岗位间的相似性，公式如下：

$$\text{sim}(j, k) = \cos(\theta) = \frac{V_j \cdot V_k}{|V_j| \cdot |V_k|} \quad (2)$$

其中， j 和 k 是岗位， V_j 和 V_k 为岗位信息特征。上式计算结果大小意味着两个岗位之间的相似程度高低，同理可以求出相似岗位集合 n^* 。

为了提升推荐的准确性和多样性，本文综合考量用户与岗位的相似度，通过加权融合的方法，计算岗位的综合得分，公式如下：

$$\text{score}(j) = \alpha * \text{sim}_u(j) + \beta * \text{sim}_i(j) \quad (3)$$

其中， $\text{sim}_u(j)$ 是基于用户的协同过滤中对岗位 j 的相似度， $\text{sim}_i(j)$ 是基于岗位的协同过滤中对岗位 j 的相似度， α 和 β 是权重系数。

利用公式(3)计算岗位综合得分值越高意味着目标更容易感兴趣，将该值降序排序，取排序较前的 N 个即TOP-N项^[12]进行推荐。

1. TF-ID方法与用户-岗位特征矩阵构建

TF-ID算法通过计算词频与逆文档频率衡量词汇的重要性，本文通过分析用户简历和职位描述的文本特征，计算匹配度，从而为用户推荐最相关的职位。考虑到用户和岗位特征的重要性，采用TF-IDF方法提取用户和岗位特征，并构建用户-岗位-特征向量。

(1) 用户-简历-特征数据集

收集用户的简历文本及岗位描述文本，构建用户-简历特征数据集。数据集中，每位用户对应一份简历，每个职位对应一份描述。

(2) 特征提取

本文使用TF-IDF方法^[13]将文本数据转换为数值特征，以量化岗位描述和用户简历中的文本信息。对于每个岗位描述和用户简历，计算每个词汇的TF-IDF值，形成特征向量。TF-IDF值的计算综合考虑了词汇在单个文档中的出现频率（词频，TF）以及该词汇在整个语料库中的逆文档频率（IDF），以此衡量词汇对于文档的重要性。

根据用户和岗位的特征向量，进一步构建用户-岗位的特征矩阵。该矩阵构建的匹配度公式：

$$\text{match}_j = \cos(\theta) = \frac{V_u \cdot W_j}{|V_u| \cdot |W_j|} \quad (4)$$

其中， V_u 表示用户 u 的特征向量，而 W_j 则表示岗位 j 的特征向量。通过公式(4)计算用户与岗位之间的匹配度，值越接近1，表示用户与岗位的匹配度越高；反之，值越接近-1，匹配度越低。用户-岗位特征矩阵如下表所示。

表1 用户-岗位特征数据集

| 用户 岗位 | j_1 | j_2 | j_3 | ... | j_n |
|---------|----------------------|----------------------|----------------------|-----|----------------------|
| u_1 | $\text{match}_{1,1}$ | $\text{match}_{1,2}$ | $\text{match}_{1,3}$ | ... | $\text{match}_{1,n}$ |
| u_2 | $\text{match}_{2,1}$ | $\text{match}_{2,2}$ | $\text{match}_{2,3}$ | ... | $\text{match}_{2,n}$ |
| u_3 | $\text{match}_{3,1}$ | $\text{match}_{3,2}$ | $\text{match}_{3,3}$ | ... | $\text{match}_{3,n}$ |
| ... | ... | ... | ... | ... | ... |
| u_m | $\text{match}_{m,1}$ | $\text{match}_{m,2}$ | $\text{match}_{m,3}$ | ... | $\text{match}_{m,n}$ |

上述矩阵体现了不同用户与各类岗位之间的匹配程度。

2. 深度学习算法

深度学习算法通过神经网络学习用户和岗位的复杂特征，从而实现更精准的推荐。本文采用AutoRec模型提取用户-岗位特征数据集进行训练和预测。

(1) AutoRec模型概述

AutoRec模型基于自编码器框架，通过编码器-解码器结构对用户-岗位特征数据进行低维特征提取与重构，适用于实际的就业推荐系统。

(2) 模型结构与工作原理

将用户对岗位的匹配度向量转换到低维的隐藏层空间中，然后输出预测值，根据预测值给用户推荐未匹配过岗位。其中，用户-岗位特征矩阵的维度与用户数 m 和岗位数 n 一致。

将用户与岗位 i 的匹配度向量 $r_{u,i} = [R_{1i}, R_{2i}, \dots, R_{mi}]$ 转换到低维的隐藏层空间中，然后输出预测值 $\hat{r}_{u,i}$ ，根据预测值给用户推荐未匹配过岗位。其中， $r_{u,i} \in R^m$ ，维度与用户数 m 一致。模型结构与

编码过程如下：

①输入层：输入用户对岗位的匹配度向量 $r_{u,i}$ 。

②隐藏层：通过编码器将 $r_{u,i}$ ，压缩为低维隐向量 h ，提取用户行为模式与岗位特征。采用单隐层设计，维度 k （典型值 $k=500$ ），权重矩阵 $V \in R^{m \times k}$ ，激活函数为 Sigmoid，隐向量 h 表示如下：

$$h = \text{Sigmoid}(V * r_{u,i} + b) \quad (5)$$

通过非线性变换将高维匹配度向量映射至低维语义空间，捕获用户偏好与岗位特征的隐含关联。

输出层：解码器基于隐向量 h 重构完整匹配度向量，填补缺失匹配度项，生成个性化推荐。权重矩阵 $W \in R^{k \times m}$ 重构预测匹配度向量：

$$\overline{r}_{u,i} = (W \cdot h + c) \quad (6)$$

输出与输入维度相同，预测未匹配项的潜在匹配度值。

AutoRec 模型的损失函数 L 通过掩码机制约束有效匹配度误差计算，并引入正则化项抑制模型复杂度，具体定义如下：

$$L = \sum_{u=1}^m \text{mask}(u) \cdot (r_{u,i} - \overline{r}_{u,i})^2 + \lambda (\|V\|_F^2 + \|W\|_F^2) \quad (7)$$

其中，通过掩码函数 $\text{Mask}(u,i)$ 实现，对权重矩阵 V 和 W 的 Frobenius 范数进行约束，默认正则化系数 $\lambda = 0.01$ 。

在 AutoRec 模型中，推荐生成阶段通过输出向量中未匹配项按预测值排序，生成 Top-N 推荐列表。

二、多算法融合的就业推荐算法

本文提出的多算法融合的就业推荐算法，通过 TF-IDF 构建矩阵以更好地捕捉用户和岗位的文本特征，并结合协同过滤算法和 AutoRec^[14] 模型进行推荐。这种融合方式整合了用户的简历信息和岗位的特征信息，能够较好地捕捉用户的个性化偏好，同时提高推荐结果的多样性和新颖性。

（一）多算法融合推荐流程

多算法融合的就业推荐算法通过以下步骤实现

（1）数据收集与预处理：收集用户简历和岗位描述的文本数据，对数据进行预处理。

（2）特征提取：使用 TF-IDF 方法将文本数据转换为数值特征，形成用户和岗位的特征向量。

（3）构建用户-岗位特征矩阵：根据用户和岗位的特征向量，构建特征矩阵，计算用户与岗位之间的匹配度。

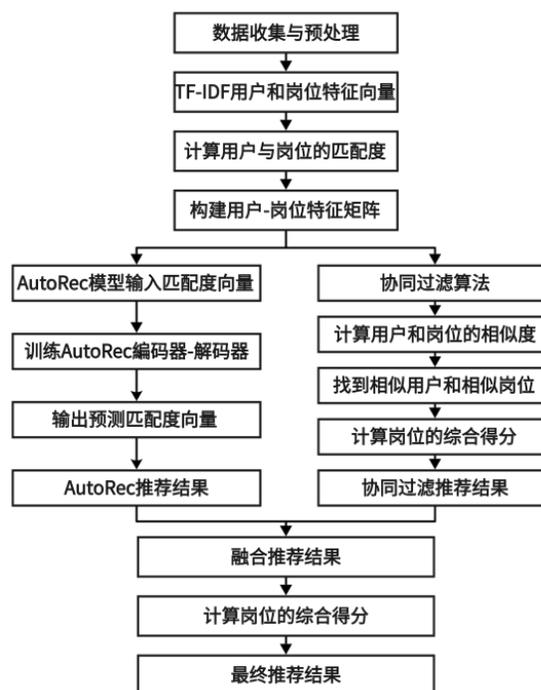
（4）AutoRec 模型应用：将用户对岗位的匹配度向量输入到 AutoRec 模型中，通过编码器-解码器结构进行低维特征提取与重构。输出预测值，根据预测值为用户推荐未匹配过的岗位。

（5）协同过滤推荐：基于用户和岗位的相似性分析，使用余弦相似度计算用户和岗位之间的相似度，并找到相似用户和相似岗位。

综合推荐结果排序：结合 AutoRec 模型和协同过滤的结果，

计算岗位的综合得分。根据综合得分对推荐岗位进行排序，生成最终的推荐列表。

多算法融合就业推荐算法的流程图如图 2 所示。



> 图2 多算法融合推荐流程

三、实验与结果分析

（一）实验环境和数据集

本文实验采用 kaggle 网站上提供的公开岗位信息和求职者数据集。其中，岗位数量 1580。包括招聘 id、招聘岗位、公司类型、薪资范围、学历、经验、工作性质、需求技能等。求职者信息数量 10891。包括用户 id、预期岗位、薪资预期、意向工作地点、期望行业、预期工作性质、工作经验、学历、是否应届、技能等。

本文中实验采用 TOP-N 方法进行推荐。实验分为训练集的 80% 和测试集的 20%。

（二）实验结果

（1）实验评价指标和参数设置

本节实验对比提出的融合算法与传统的协同过滤的推荐效果，使用 MAE、RMSE 作为度量标准。

在协同过滤算法中，需确定与目标用户最为相似的用户数量 $n=35$ ，经实验验证^[15]；在推荐列表 $N=10,15,20,25,30$ 条件的情况下，同时确定协同过滤推荐算法的权重 0.5，在 AutoRec 模型中，其关键参数维度取值设定 $k = 500$ 。

（2）实验对比融合算法、协同过滤算法。使用 MAE 作为度量标准验证两种算法的推荐效果。根据预测匹配度，为每个用户生成 Top - N 推荐列表，实验如表 2，表 3 所示。

表2 不同算法实验数据 MAE指标

| 算法 | TOP-N | 10 | 15 | 20 | 25 | 30 |
|------|-------|-------|-------|-------|-------|-------|
| 协同过滤 | MAE值 | 0.771 | 0.785 | 0.768 | 0.767 | 0.763 |
| 融合算法 | MAE值 | 0.768 | 0.779 | 0.768 | 0.764 | 0.762 |

表3 不同算法实验数据 RMAE指标

| 算法 | TOP-N | 10 | 15 | 20 | 25 | 30 |
|------|-------|-------|-------|-------|-------|-------|
| 协同过滤 | RMAE值 | 0.997 | 0.995 | 0.982 | 0.978 | 0.977 |
| 融合算法 | RMAE值 | 0.989 | 0.981 | 0.975 | 0.974 | 0.974 |

通过实验对比上述两种算法的 MAE、RMAE 值，融合算法在 N 取不同的值时的 MAE 值与协同过滤算法相近或略低，表明融合算法具有更好的推荐准确性。

四、结语

通过实验可以得出：本文提出的多算法融合的就就业推荐算法，充分利用用户和岗位的特征信息，能够有效提高推荐的精确度。但在研究中，对于用户不同时期的特征偏好，以及外部原因的影响，如岗位热度等，使得用户对某一岗位特征偏好的改变不能够进行及时修正，下一步将进行完善研究。

参考文献

- [1] 薛亮, 冯尊磊, 凌兴宏, 等. 基于标签聚类和协同过滤算法的就就业推荐系统设计 [J]. 微型电脑应用, 2024, 40(01): 1-4.
- [2] 韩义畴, 郑香山, 王广铭, 等. 高职院校学生就业能力现状分析与培育对策 [J]. 现代商贸工业, 2024, 45(06): 109-112.
- [3] 蒋大锐, 徐胜超. 基于模糊层次聚类的大学生就业数据分类存储系统 [J]. 现代电子技术, 2024, 47(03): 123-129.
- [4] 李龙, 金铄, 黄霞. 基于改进 TF-IDF 算法的毕业生就就业推荐算法研究 [J]. 计算机与数字工程, 2023, 51(09): 1985-1989+2118.
- [5] 崔秀艳. 基于大数据分析的人工智能的智慧就就业推荐系统设计及实现 [J]. 河北软件职业技术学院学报, 2023, 25(02): 15-19.
- [6] 江宜橙. 基于深度学习的高校学生就就业推荐系统研究与应用 [D]. 哈尔滨工程大学, 2022.
- [7] 金铄. 基于用户画像的高校毕业生就就业推荐方法研究 [D]. 东北石油大学, 2022.
- [8] Wan P. Development of the Employment Recommendation System based on K-Means Improved Collaborative Filtering Algorithm [J]. 2022.
- [9] Huang S, Huang X, Zeng T, et al. A long command subsequence algorithm for manufacturing industry recommendation systems with similarity connection technology [J]. Applied Mathematics and Nonlinear Sciences, 2022.
- [10] 付洋, 周勇, 马凯. 基于协同过滤的大学生就就业推荐算法研究 [J]. 黑龙江科学, 2024, 15(16): 90-93.
- [11] 张小雷. 协同过滤算法改进实验及对比分析 [J]. 电脑知识与技术, 2022, 18(35): 64-66. DOI: 10.14004/j.cnki.ckt.2022.2147.
- [12] 黄俊萍. 协同过滤算法在大学生就就业推荐系统中的应用 [J]. 信息技术与信息化, 2023, (12): 93-97.
- [13] 林振荣, 黄虹霞, 舒伟红, 等. 基于 TF-IDF 与用户聚类的推荐算法 [J]. 计算机仿真, 2022, 39(06): 341-345.
- [14] 严武军, 刘守业, 贺娇娇. 基于 AutoRec 的推荐系统模型研究 [J]. 现代计算机, 2023, 29(24): 69-73.
- [15] 黄虹霞. 基于频繁项集挖掘和用户聚类的协同过滤算法研究 [D]. 南昌大学, 2021.