

基于情境感知和协同过滤的就业推荐算法研究

焦念莱¹, 付晓², 朱琳³, 陈永磊^{2,*}

1. 计算机科学与工程学院, 山东科技大学, 山东青岛

2. 海洋科学与工程学院, 山东科技大学, 山东青岛

3. 青岛市保密科技测评中心, 山东青岛

摘要: 针对全球经济衰退, 企业发展受阻, 裁员缩招, 高校毕业生就业越来越严峻的问题, 提出一种基于情境感知和协同过滤的就业推荐算法, 帮助高校毕业生更加及时、准确的获取高质量、有针对性的招聘信息。本文对传统协同算法和情境融合算法进行介绍, 并对基于情景感知和用户的协同过滤就业推荐算法进行分析介绍, 使用自定义数据集进行实验, 并与传统协同过滤算法比较分析结果, 证明本文算法的有效性, 最后对本文提出的算法进行总结。

关键词: 情境感知; 协同过滤; 高校大学生; 就业推荐

Research on Employment Recommendation Algorithm Based on Context-Awareness and Collaborative Filtering

Nianlai Jiao, Xiao Fu, Lin Zhu, Yonglei Chen

1. College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong

2. College of Marine Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong

3. Qingdao Confidentiality Technology Evaluation Center, Qingdao, Shandong

Abstract: In response to the global economic recession, the development of enterprises is hampered, layoffs and downsizing, and the employment of college graduates is becoming more and more severe, a context-aware and collaborative filtering-based employment recommendation algorithm is proposed to help college graduates obtain high-quality and targeted recruitment information more timely and accurately. In this paper, we introduce the traditional collaborative algorithm and the context fusion algorithm, and analyze the collaborative filtering employment recommendation algorithm based on context-awareness and users. We conduct experiments using self-defined data sets and compare the results with the traditional collaborative filtering algorithm to prove the effectiveness of the algorithm in this paper.

Keywords: Context-awareness; Collaborative filtering; College students; Employment recommendation

* 作者简介: 焦念莱(1994-)男, 山东青岛人, 山东科技大学计算机科学与技术学院辅导员, 助教, 从事大学生思想政治研究; 付晓(1979-), 女, 山东青岛人, 山东科技大学海洋科学与工程学院组织员, 副教授, 从事大学生思想政治研究; 朱琳(1998-), 女, 山东滨州人, 青岛市保密科技测评中心助理工程师, 从事信息系统测评研究; 陈永磊, 通讯作者, 山东科技大学海洋科学与工程学院副院长, 副教授, 从事大学生思想与政治研究。

1 研究背景

从2025届全国普通高校毕业生就业创业工作会议获悉：2025届高校毕业生规模预计达1222万人，同比增加43万人，无论在规模还是数量上都创下了历史新高。而国家统计局2022年11月份的数据显示[1]，16-24岁劳动力调查失业率为17.1%。在就业压力巨大、竞争如此激烈的情况下，利用好“互联网+就业”模式，对大学生进行就业指导规划显得尤为重要。在大数据背景下，推荐系统逐渐被很多行业使用，其中，基于协同过滤的算法[2]，是目前最成熟、应用最广泛的一种。本文结合学生用户群体实际情况与影响就业选择的因素，提出了一种基于情境感知和协同过滤的就业推荐算法，根据学生不同的个人情况进行精准且符合个性化需求的就业推荐。

2 传统协同过滤推荐算法介绍

推荐算法是计算机与数学相结合的技术，其作用是根据用户或者他人的行为数据，根据数学公式，为用户推荐感兴趣或者符合自身需求的内容。目前个性化推荐算法分为5种，其中协同过滤具有模型通用性强、效果好等优点，是目前个性化推荐算法研究和应用最广泛的推荐算法。协同过滤（Collaborative Filtering, CF）推荐方法[2]的主要思想是，利用已有用户群的历史行为数据，计算有共同项目的用户间偏好的相似性，预测当前用户最可能感兴趣的事物。CF将用户评分作为唯一的知识源，不需要用户的个性特征和物品的属性信息。传统的协同过滤分为基于用户的协同过滤和基于项目的协同过滤。

2.1 基于用户的协同过滤

基于用户的协同过滤主要考虑用户之间的相似度。首先获得用户-项目评分矩阵，如表1所示。通过用户-项目评分矩阵计算用户之间的相似度，找出与目标用户相似度最高的前N个用户作为邻居用户，根据N个邻居用户的已评分物品预测目标用户对相应物品的评分，找到评分最高的Top-K个物品推荐给目标用户。基于用户的协同过滤算法原理图

如图1所示。

基于用户的协同过滤算法[3]由于需要寻找计算目标用户与其他用户之间的相似性关系，计算复杂度较于基于项目的协同过滤算法更高。但是不同用户喜欢的领域相差较大，该算法能够帮助用户发现其他领域的惊喜物品，拓宽兴趣范围。

表1. 项目-评分矩阵

项目 用户	i_1	i_2	i_3	i_4	...	i_n
u_1	2	3	?	1	...	?
u_2	3	?	2	?	...	3
u_3	?	1	1	?	...	2
...	...	...	...	...	...	...
u_n	2	4	?	4	...	?

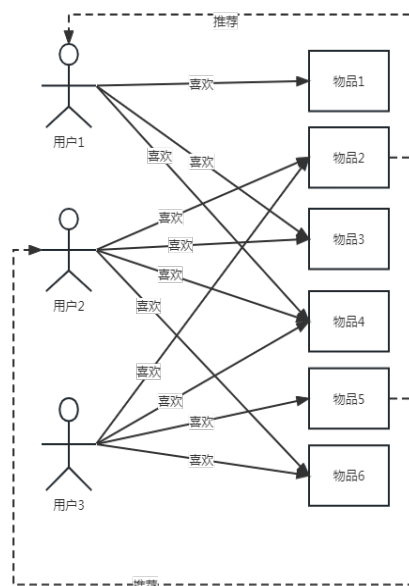


图1. 基于用户的协同过滤算法原理图

2.2 基于项目的协同过滤

基于项目的协同过滤算法[4]主要考虑项目之间的相似度。根据所有用户对其有关项目的评分矩阵计算项目间的相似度得到相似性矩阵。根据目标用户的有关项目选取相似值最高前N个项目进行评分预测，排序得到评分最高的前Top-K个物品推荐给目标用户。基于项目的协同过滤算法原理图如图2

所示：

基于项目的协同过滤算法适用于项目数量明显少于用户数的情况，利用用户的历史行为作推荐解释，更能够使用户信服。但需要及时离线更新项目相似度矩阵后，才能够推荐新项目给用户。

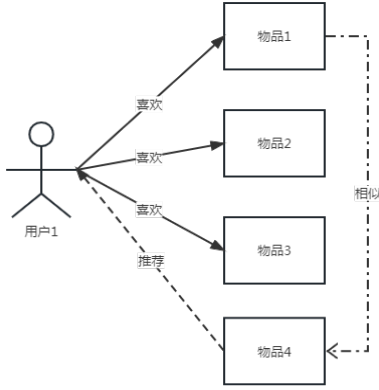


图2. 基于项目的协同过滤算法原理图

2.3 协同过滤相似性计算

计算用户或项目之间的相似性是协同过滤中重要环节，通过相似性计算来选取待研究的群体，计算方法主要有杰卡德相似[5]、余弦相似[6]、皮尔逊相关[7]等。余弦相似性是用两个n维向量之间的夹角来测算相似度，若两个物品a和b用对应的评分向量 $\vec{a}$ 、 $\vec{b}$ 表示，则其相似度可用公式（1）计算：

$$\text{sim}(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| * |\vec{b}|} \quad (1)$$

其中，符号 $\cdot$ 表示两个向量之间的点积， $|\vec{a}|$ 表示向量的模， $\text{sim}(\vec{a}, \vec{b})$ 表示相似度。相似度的值介于0和1之间，越接近1表示越相似。但是此方法并没有考虑用户打分尺度的问题，如有的用户喜欢打高分，有的用户偏好打低分。因此对余弦相似性进行改进，改进版的方法是在评分值中减去均值。改进版公式为：

$$\text{sim}(a, b) = \frac{\sum_{u \in U} (r_{u,a} - \bar{r}_u)(r_{u,b} - \bar{r}_u)}{\sqrt{\sum_{u \in U} (r_{u,a} - \bar{r}_u)^2} \sqrt{\sum_{u \in U} (r_{u,b} - \bar{r}_u)^2}} \quad (2)$$

其中，U为所有同时给项目a和b评分的用户集， $\bar{r}_u$ 为用户u评分的均值， $r_{u,a}$ 为用户u对项目a的评分， $r_{u,b}$ 为用户u对项目b的评分。相应地，改进

后的余弦相似性取值范围是-1到+1之间。

3 情境感知介绍

一般来说，用户的学历层次、性别、时间段、就业地点、薪资要求等情境信息均是影响就业选择的因素。有些情境信息是经常变化的，系统只有动态地获取用户当前的情境，才能更准确地为用户推荐感兴趣的工作。协同过滤只以评分作为标准，忽略了用户的兴趣、环境等因素发生的变化。因此，本文通过融合情境感知[8]对协同过滤算法进行改进，解决当前协同过滤算法的不足。基于情景感知的协同过滤算法问题描述为：用户集合U，工作集合P，情境因素集合E，用户对工作的评分集合I。算法首先将“项目-评分”矩阵和“项目-属性”矩阵相结合，建立“用户-项目-情境”矩阵。根据“用户-项目-情境”矩阵信息，选择适合该资源推荐的情境因素，将情境因素与工作类型进行关联，计算不同情境条件下的情境熵和情境权重，来衡量不同情境因素对工作选择的重要程度。

3.1 项目-评分-情景矩阵构建

每位用户在挑选感兴趣的工作时都会包含一个情境主题，情境主题中包含用户各个情境因素的特征，采用空间向量表示用户当前的情境主题信息。通过调查分析，与大学生找工作相关的情境因素主要包括学历层次、性别、薪资要求、期望就业类型、期望就业地区等，具体分析如表2所示。

工作即为表1中的项目，将工作-评分矩阵与工作属性结合起来，构建用户-情境-评分矩阵 $R(U, E, I) = \text{user} * \text{context} * \text{resource}$ 。如表3所示，表中数字表示某位用户的所有情境属性值， U_j 表示用户j， I_{ji} 表示用户j在情境 E_i 下对工作i的评分， I_{type} 为工作的类型。

3.2 情景感知计算

3.2.1 情境条件熵的计算

在就业推荐模型中，不同情境因素对推荐结果的影响度不同。比如学历层次对工作选择的影响要远大于英语水平因素产生的影响。因此，在工作推

表2. 大学生找工作时的情境因素

情境因素	情境属性值	情境表示
学历层次	大专、本科、硕士、博士	大专=1; 本科=2; 硕士=3; 博士=4
性别	男、女	男=1; 女=2
学科门类	哲学、经济学、法学、教育学、文学、历史学、理学、工学、农学、医学、军事学、管理学、艺术学	哲学=1; 经济学=2; 法学=3; 教育学=4; 文学=5; 历史学=6; 理学=7; 工学=8; 农学=9; 医学=10; 军事学=11; 管理学=12; 艺术学=13
薪资要求	3000以下、3000-5000、5000-8000、8000-10000、10000-13000、13000以上	3000以下=1; 3000-5000=2; 5000-8000=3; 8000-10000=4; 10000-13000=5; 13000以上=6
英语水平	优秀、良好、合格	优秀=1; 良好=2; 合格=3
计算机水平	优秀、良好、合格	优秀=1; 良好=2; 合格=3
学生干部任职情况	主要学生干部、普通学生干部、无任职经历	主要学生干部=1; 普通学生干部=2; 无任职经历=3
期望就业类型	公职、大型企业、中小企业、创业公司	公职=1; 大型企业=2; 中小企业=3; 创业公司=4
期望就业地区	东部沿海地区、中部内陆地区、西部边远地区、生源地区	东部沿海地区=1; 中部内陆地区=2; 西部边远地区=3; 生源地区=4
性格	胆汁质、多血质、粘液质、抑郁质	胆汁质=1; 多血质=2; 粘液质=3; 抑郁质=4

表3. 用户-情境-评分数据表

用户	E_1	E_2	E_3	E_4	I_{ji}	I_{type}
U_1	1	1	2	3	I_{1i}	I_{type1}
U_2	3	2	1	4	I_{2i}	I_{type2}
U_3	1	1	3	2	I_{3i}	I_{type3}
...	...	...	...	...	...	...
U_n	2	2	2	6	I_{ni}	I_{typen}

荐过程中, 影响力较大的因素在模型中占的权重也应该更大。情境因素对推荐结果的影响度可以用条件熵[9]来表示。信息熵主要用来度量随机变量的不确定性, 随机变量的取值个数越多, 混乱程度则越大, 信息熵越大, 随机变量的不确定性越大。随机变量X的信息熵定义为 $H(X)$, 计算公式为:

$$H(X) = -\sum_x p(x) \log p(x) \\ = -\sum_{i=1}^n p(x_i) \log p(x_i) \quad (3)$$

其中, $p(x_i)$ 表示取值为 x_i 的概率。

条件熵表示在完全已知一个变量的基础上, 另一个变量的不确定程度。表示在已知随机变量X的条件下随机变量Y的不确定性, 计算公式如下:

$$H(Y|X) = \sum_x p(x) H(Y|X=x) \\ = -\sum_{x,y} p(x,y) \log p(y|x) \quad (4)$$

其中, $p(y|x)$ 指在随机变量x发生的情况下y取值的概率。当 $p(y|x) = 0$ 时, 表示在情境条件x的条件下时, 不会去选择资源y; 当 $p(y|x) = 1$ 时, 则表示当情境条件x下一定会选择资源y, 此时情境条件熵 $H(Y|X) = 0$, 即情境条件x对选择资源y的影响最大。

3.4.2 情境权重的计算

情境权重用于表示不同情境因素对工作选择的重要程度。在某种情境下, 情境条件熵越小, 则在此情境下选择工作的不确定性程度越小, 此情境因素对工作的影响程度越大, 情境权重越大。情境权重的计算公式如下:

$$W_x = \frac{1-H(Y|X)}{N-\sum_{x \in X} H(Y|X)} \quad (5)$$

其中N表示情境属性的个数, $H(Y|X)$ 表示X情

境条件熵的值。

4 基于情景感知和用户的协同过滤就业推荐算法

4.1 构建用户相似度模型

根据用户在当前情景下的行为，分析与其相似的用户，得到目标用户与其相似用户共同项目的评分矩阵。选用皮尔逊相关系数[7]对评分矩阵进行相似性计算，得到目标用户的前K个近邻用户。设目标用户a，相似用户b，共同项目评分矩阵为P， $r_{a,p,c}$ 和 $r_{b,p,c}$ 分别表示用户a和b在情景c下对工作p的评分，和分别表示用户a和用户b在情景c下对所有共同项目的平均评分，根据公式（6）计算用户a和用户b的相似性 $\text{sim}(a,b)$ 。

$$\text{sim}(a,b) = \frac{\sum_{p \in P} (r_{a,p,c} - \bar{r}_{a,c})(r_{b,p,c} - \bar{r}_{b,c})}{\sqrt{\sum_{p \in P} (r_{a,p,c} - \bar{r}_{a,c})^2} \sqrt{\sum_{p \in P} (r_{b,p,c} - \bar{r}_{b,c})^2}} \quad (6)$$

4.2 预测评分计算

根据4.1中计算得到的近邻用户集合，使用公式（7）分别计算所有近邻用户相关项目在情景c下针对用户a的预测评分 $\text{pred}(a,p,c)$ 。

$$\text{pred}(a,p,c) = \bar{r}_{a,c} + \frac{\sum_{b \in N} \text{sim}(a,b) * (r_{b,p,c} - \bar{r}_{b,c})}{\sum_{b \in N} \text{sim}(a,b)} \quad (7)$$

根据公式（5）计算得到的情境权重和得到的预测评分，使用公式（8）计算在情境c下用户a对项目p的最终预测评分。

$$r_{a,p,c} = \sum_{E_i \in E} W_{c,E_i} * \text{pred}(a,p,c) \quad (8)$$

其中， W_{c,E_i} 表示在情景c下情景要素 E_i 的权重。

基于情景感知和用户的协同过滤就业推荐算法流程描述如下：

输入：用户集合U，工作集合P，用户-情境-评分矩阵 $R(U,E,I)$ 。

输出：用户对工作的预测评分V。

1) 对输入数据进行初步处理，得到用户总数n，工作总数m，情境因素总数k等。

2) 建立“用户-项目”矩阵、“项目-评分”矩阵、“项目-属性”矩阵及“用户-项目-情境”

矩阵等。

3) 根据公式（6）构建用户相似度模型。

4) 根据公式（7）和公式（8）计算工作的预测评分。

5) 对计算得到的工作评分预测进行过滤去重。

6) 根据最终的预测评分排序生成目标用户的推荐工作列表。

5 数据集与实验

5.1 数据集

本文以中国某高校2020届、2021届5000多名学生的就业数据为实验数据集。首先通过学生信息数据库获取包括学生的性别、学历层次、生源地、专业、学生干部任职情况等在内的基本信息；其次通过追溯毕业生情况获取了该部分学生的工作情况，主要包括企业名称、就业类型、企业所在地、工作满意度等；最后通过问卷调查形式获取学生对列出的样本工作的兴趣评分作为“项目-评分”评分矩阵数据。

5.2 评价指标

本文选用覆盖率（Coverage）、准确率（Precision）两个评价指标对本文提出的方法和经典协同过滤推荐算法进行对比分析。覆盖率指推荐系统可以推荐出的项目占总项目集合的比例，能够反映推荐系统的发掘能力。覆盖率的计算公式如公式（9）：

$$\text{Coverage} = \frac{|U_{u \in U} R(u)|}{|I|} \quad (9)$$

其中，U为用户集合，I为项目集合，推荐给每个用户的项目列表为 $R(u)$ 。覆盖率越接近1说明算法推荐结果涉及的项目范围越大，推荐效果越好。

准确率是指推荐算法推荐给目标用户的项目列表占用户实际选择项目的比例。准确率定义如公式（10）所示。

$$\text{Precision} = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|} \quad (10)$$

5.3 实验结果及分析

为验证和分析本文提出的推荐算法的效果，使用5.1介绍的数据集，75%为训练集，25%为测试集，进行本文算法与以下两种算法的对比实验：

1) 传统的基于用户的协同过滤推荐算法 (UserCF)。

2) 传统的基于项目的协同过滤推荐算法 (ItemCF)。

5.3.1 探究不同K取值下准确率、召回率

三种算法在不同K取值下的准确率如图3所示。

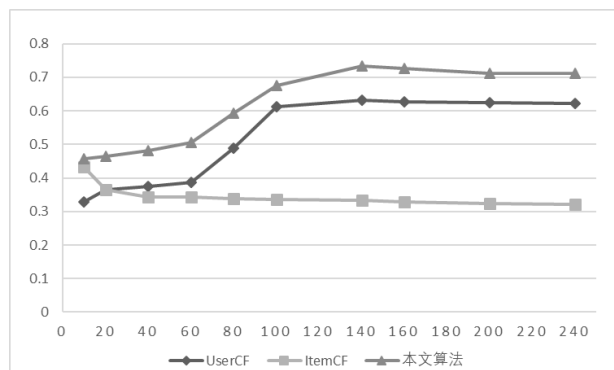


图3. K取不同值时的准确率

分析图3可得出如下结论：在三种算法中，本文提出的算法准确率最高，最优值为73.26%。在K取值小于140时，三种算法的准确率均随着K值的增大而提高。在K取值为140时，基本达到了最优值，随着K取值的继续增加，准确率趋于平稳。在K取值较小的情况下，基于项目的协同过滤算法的准确率要高于基于用户的协同过滤算法，随着K值的增大，准确率呈下降趋势。

5.3.2 覆盖率对比

当K取值为140时，三种算法的覆盖率如表4所示。

表4. 当K取值为140时，三种算法的覆盖率	
算法	覆盖率 (%)
UserCF	65.3%
ItemCF	54.7%
本文算法	78.2%

由表4内容可知，传统的基于用户的协同过滤

算法和基于内容的协同过滤算法覆盖率均较低，分别为65.3%和54.7%。而本文提出的融入情境感知的协同过滤算法覆盖率达到78.2%，远高于两种传统的协同过滤算法。覆盖率的提高说明在算法种融入情境感知因素后，能够针对当前情境为目标用户构建更符合用户的喜好模型，使得更多的用户可以得到与自身行为喜好相关的更加满意的推荐结果。

6 结语

随着全球经济的发展减缓，企业发展受阻，裁员缩招，就业形势越来越严峻，而我国的高校毕业生数量居高不下，招聘市场的竞争也越来越激烈，求职者们需要更加高效和准确地找到适合自己的工作。为此，许多研究者提出了各种不同的就业推荐算法，以帮助求职者更好地进行职业规划和择业。

本文主要研究了基于情境感知和协同过滤的就业推荐算法。首先，基于用户的协同过滤算法可以根据求职者的兴趣爱好、薪资要求等信息，推荐其他具有相似画像求职者感兴趣的工作，并通过模型计算这些工作的预测评分，选择评分最高的一部分工作推荐给该求职者。其次，通过对求职者的个人信息、当前状态等情境因素的分析，可以自动筛选出当前情境下最合适的工作推荐，可以提高推荐的准确率，提高用户的满意度。

该算法的优点在于，通过情境感知和协同过滤的相互结合，可以更加精准地为求职者提供个性化的推荐服务。然而，该算法也存在一些不足之处。首先，求职者提供的信息可能不够准确或完整，从而影响推荐结果的准确性。其次，该算法还需要不断收集和更新求职者的信息，才能保持其推荐的有效性。

综上所述，基于情境感知和协同过滤的就业推荐算法是一个非常有潜力的研究方向。通过不断优化算法，我们可以为高校毕业生甚至更多的求职者提供更加准确和个性化的推荐服务，帮助他们更及时、准确获取高质量、有针对性的招聘信息。

参考文献

[1]国家统计局[EB].<http://www.stats.gov.cn/>,2022.11.30.

- [2]赵俊逸,庄福振,敖翔等.协同过滤推荐系统综述[J].信息安全学报, 2021,6(05):17-34.
- [3]肖梅.基于用户属性特征与评分的协同过滤算法[J].现代信息科技, 2022,6(21):91-93.
- [4]马瑞敏.基于项目的协同过滤推荐算法研究[J].晋中学院学报, 2021,38(03):40-43.
- [5]周朝阳.杰卡德相似度在图书推荐中的应用研究[J].情报探索,2017(07):43-46.
- [6]李一野,邓浩江.基于改进余弦相似度的协同过滤推荐算法[J].计算机与现代化,2020,293(01):69-74.
- [7]彭海.皮尔逊相关系数应用于医学信号相关度测量[J].电子世界, 2017,07:163.
- [8]张磊.基于情景感知技术的网络信息资源个性化推荐方法[J].河北北方学院学报(自然科学版),2022,38(01):21-26.
- [9]王茜,张贤勇,吕智颖.不完备决策信息系统的混合条件熵与多属性决策[J].系统工程理论与实践,2022,42(12):3401-3411.

