

基于软集理论驱动的毕业生推荐系统研究

杨佳恂

(郑州航空工业管理学院 信息科学学院, 郑州 450015)

摘要:网络技术的发展为毕业生与企业之间的顺利沟通提供了支持,但大多数招聘或推荐平台依然是各自“表演”的模式,缺乏有效对接.面向毕业生与企业的直接需求、核心需求,在软集理论驱动的基础上,构建一个智能的、支持决策的毕业生推荐平台.该平台允许用户自主设置属性及属性排序,对推荐结果进行优劣排序,使得推荐结果更加客观全面,符合大学生及企业的需求.最后实验表明该方法具有较高的性能.

关键词:智能推荐;软集;决策;毕业生推荐;平台

中图分类号:TP319

文献标志码:A

1 研究背景

近些年,随着国家对高等教育的关注,教育意识得到普及,每年都有众多的高校毕业生涌向就业市场.然而,“史上最难就业季”、“史上最难招聘季”却频频发生.根据有关资料显示,大量的毕业生在寻找工作的时候往往毫无重点和目标,他们需要花费大量的时间和精力在人才市场、智联招聘、58同城等渠道上投放自己的个人简历,在各个用人单位之间流连择选,最后却常常无功而返.另一方面,很多用人单位常年处于员工不足,急招急聘的状态.除此之外,即便毕业生和用人单位之间达到暂时的妥协,签订了劳务合同.毕业生对于公司、工作的满意度却很低,很难愉快地工作,而用人单位对员工的表现也处于不满之中,期望寻找更适合企业发展的优秀员工,两者之间存在着本质需求的差异^[1].这些现象之所以产生并不是毕业生学识不够、专业度太差,更不是就业市场困难、供大于求.而是,毕业生与用人单位之间无法协调,没有任何一个平台能够将毕业生和用人单位相互连接,充分综合两者的需求,在其中找到一个合适的平衡点,既可以满足毕业生对于所学专业、就业倾向、就业地点、薪资等要求,也可以满足用人单位对于员工学历、职业素质、工作经验等要求.鉴于当下的就业现状,以及计算机、互联网的快速普及,通过先进的技术手段和科学理论,研发出一个针对毕业生就业的推荐系统成为当下的需求,也是完全可以实现的一个创意.

以往的就业信息系统侧重于生搬硬套,将毕业生和企事业单位固定在一个清晰的框架中,无法针对具体的需求和情况灵活多变,所以已经无法满足当今多变的就业市场.本文根据实际情况研发了基于软集合理论的毕业生推荐系统.在互联网上充斥着大量的信息,涵盖了员工的基本情况,该系统可以通过数据挖掘技术,扫描互联网上的可靠信息,形成对某一毕业生用户的整体认识,结合用人单位的需求,对该毕业生用户进行合理的推荐^[2].推荐系统立足于先进的模糊数学与软集理论设计,具有很强的专业性和市场前景,能够解决当下就业的社会难题,已经成为很多专家学者研究的对象.

2 理论支持

在以往的工作中,存在着一些暂时无法确定的问题,为了解决这些处于变化中无法清晰界定的现象,学

收稿日期:2015-04-17;修回日期:2015-06-13.

基金项目:河南省科技攻关项目(112300410304);郑州航院2014年教育科学研究基金项目(zhjy15-24);河南省大中专毕业生就业创业课题(JYB2013029).

作者简介(通信作者):杨佳恂(1978-),女,河南郑州人,郑州航空工业管理学院讲师,主要从事面向就业的信息技术研究,E-mail:yjy@zzia.edu.cn.

者提供了一系列专业理论.例如,概率论、模糊集理论、区间数学、粗糙集等,然后这些理论由于自身参数不够完善、功能存在缺陷,无法更好地为科学研究提供帮助.相比之下,著名学者 D. Molodtsov 的软集合理论更为完善,对解决这些变动中的问题具有重要的意义,因此利用更为广泛.之所以在设计就业推荐系统的时候考虑到软集合理论,是因为之前由 P. K. Maji 等学者设计的决策系统,正因为恰当地结合了软集合理论,取得了系统试用的优质效果^[3].因此,设计的类似于决策系统的就业推荐系统也非常需要软集合理论的支持.

针对软集合的理论及分类方法,本文依据相关文献列出 5 种定义.

定义 1 (F, E) 是论域上的一个软集合(*soft set*)当且仅当 F 是 E 到论域 U 的所有子集的一个映射.也就是说,设定 U 是初始论域, E 是一个参数集,其中 $P(U)$ 是 U 的幂集,当且仅当 F 是 E 到 U 的所有子集的一个映射时,称为 (F, E) 是 U 上的一个软集合.该参数族中每个集合 $F(\epsilon)$ ($\epsilon \in E$) 是 (F, E) 的 ϵ 个元素的集合,或者是 (F, E) 中 ϵ 个相似元素的集合^[4].

定义 2 设有一个二元运算符 $*$,其中 (F, A) 和 (G, B) 是论域 U 上的 2 个软集,则在软集上的运行 $*$ 可表示为 $(F, A) * (G, B) = (H, A \times B)$,其中 $(\alpha, \beta) = F(\alpha) * G(\beta)$, $\alpha \in A, \beta \in B, A \times B$ 是集合 A 和集合 B 的笛卡尔集.

定义 3 设 $P(U)$ 为论域 U 上所有的模糊集, E 是它的一个参数集,其中 $A_i \subset E$,集合对 (F_i, A_i) 被称为论域 U 上的一个模糊软集,当且仅当 F_i 是 A_i 到 $P(U)$ 的一个映射,如 $F_i: A_i \rightarrow P(U)$ ^[5].

定义 4 设信息系统可以表示为 $S = (U, A)$,其中 U 是对象的非空有限集合; A 为属性的非空有限集合,对于任意满足的 $a \in A$ 有 $\alpha: U \rightarrow V_a, V_a$ 为属性 α 的值域^[6].

定义 5 若存在 E 的子集 $A = \{e_1, e_2, \dots, e_p\}$ 满足 $\sum_{e_k \in A} h_{1k} = \sum_{e_k \in A} h_{2k} = \dots = \sum_{e_k \in A} h_{nk}$, 则 A 不是必要的,否则 A 是必要的.若 B 是必要的且 $\sum_{e_k \in E-B} h_{1k} = \sum_{e_k \in E-B} h_{2k} = \dots = \sum_{e_k \in E-B} h_{nk}$, 则 B 为 E 的一个正则化约简, $E-B$ 是满足 $f_{E-B}(\cdot)$ 为一常数 E 的最大子集^[7].

3 毕业生推荐系统设计

3.1 推荐系统功能模块

对于绝大部分的推荐系统来说采用的结构模式都是 B/S 的样式,本文所研究的毕业生推荐系统也是采用这样的架构样式的,从全局来看这个系统是可以分为两个部分分别是后台和前台,所谓的后台就是数据处理和搜索引擎,而前台则是指该系统与使用者交流的界面,前后台之间是通过数据库连接的,图 1 就是这个系统的功能模块简略说明.

从图 1 中,可以到该系统具备以下功能.

1) 在系统的操作界面部分,用户可以看到推荐的列表,这个列表是在用户提出需求的情况下,由系统自动筛选最适应该用户的结果,提供给用户.用户看到推荐列表后,可以根据自己的需要排序.用户需要包括求职意向、专业、岗位等.

2) 那些用于解决不确定问题的集合被称之为软集合,本系统中的软集量化模块就是主要针对这样的问题,比如说在系统中的那些求职、招聘信息可能存在多余的字段,所以这个模块的存在就是为了减少这些冗余信息的存在,提高系统的运算时间,而且还可以将这些作为客户端信息输入的一个标准^[8].

3) 系统软集决算是系统中的一个核心的部分,主要的作用是为了给招聘和应聘的双方可以更快更好地找到匹配的信息,该算法实现逻辑后文介绍.

4) 多属性权重计算组件主要面向招聘信息和求职信息的计算处理,负责权重计算.

不同的工作侧重的点是不同的,所以各自侧重的权重也是不一样的.为每个需要招聘的企业或者公司以

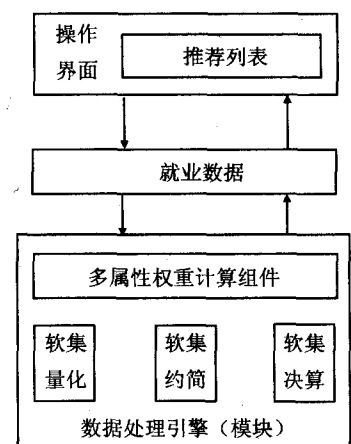


图1 功能模块简略说明

及找工作者提供准确核实的参考信息是本系统的最终目的。

3.2 前台的工作过程

对于一个企业来说,主要是希望通过网页来实现对于各类型的毕业生的求职需求,同时也需要对于毕业生进行筛选和分类并且赋予不同的关注度.在服务器端,当其在接收到了查询的要求后会调用与之对应的引擎来进行一个计算和处理^[9].当这一流程完成后会从数据库中进行筛选同时将有关的信息数据进行提取展示,用户端也可以通过显示内容寻求的自己所需信息.

3.3 后台的工作过程

后台的工作核心就是引擎对于数据信息的处理,其中主要包含的一个模块就是多属性的决策组件,具体的工作流程如下.比如说,当毕业生或者企业公司的招聘人员在使用查询的功能时,该系统就自动将数据库中的相关求职者的所有信息按照一定的排列顺序,经过一定的数字决算之后,将数据信息显示在查询用户的客户端上,在显示的数据中就可以发现推荐的那些学生中的各项信息的排名情况,用户可以根据自己的需要将有用的信息存储起来^[10].当然这个系统也同样是作用于就业的大学生们,就业者们可以根据自己的要求,在搜索栏中输入相应的命令,引擎会根据之前企业和公司之前录入在数据库中的信息进行比对和分析,通过相应的多属性的决策分析后将核实的信息显示在求职毕业生的查询结果的页面中.

4 人才推荐平台中基于软集理论的核心技术

4.1 优化模糊软集的权重

设论域 U 的所有模糊集为 $P(U)$, E 为参数集, $A \subset E$, 三元组 (F, A, ω) 被称为 U 上一个带权模糊软集, 其中集合 (F, A) 为 U 上的一个模糊软集, $\omega: A \rightarrow [0, 1]$ 为权重, 因此任意 $A_i \in A$ 它的权重为 $\omega_i = \omega(A_i)$.

通过优化带权重模糊软集, 能够做到识别信息系统中不同要素的重要程度, 通过权重函数 ω 来指定每个要素的重要性^[11].

4.2 构造系统的软集信息表

本文以人才推荐平台软集信息表进行说明, 系统中表示为 $U = \{o_1, o_2, \dots, o_n\}$, 用人单位提出的人才要求指标和待遇参数分别表示为 $E = \{e_1, e_2, \dots, e_m\}$, 设定 $\omega = \{\omega_1, \omega_2, \dots, \omega_m\}$ 为用人单位对各项指标的重视程度, 即 $\omega_1 + \omega_2 + \dots + \omega_m = 1$, 其中 $e_{ij} \in [0, 1]$, 当 $e_{ij} = 0$ 时表示应聘者在属性 e_i 上完全不符合用人单位的需求, 当 $e_{ij} = 1$ 时表示, 用人单位在属性 e_j 上完全符合应聘者的需求^[12]. 通过实例分析, 所得信息表如表 1 所示.

4.3 软集 (F, E) 的对照表构造方法

软集 (F, E) 的对照表由 n 行 n 列构成, 含 n 个对象, 二维表中每一元素标记为 C_{ij} , 值为在软集 (F, E) 中对象 O_i 的取值 $\geq O_j$ 值所对应的参数个数. O_i 在 e_i 上的取值必须参考属性的权重. 如在表 1 中 o_1 在 e_1 上的取值应为 $\omega_1 \times 0.5 = 0.15$. 则有 $0 \leq C_{ij} \leq d$ 且 $C_{ij} = d, \forall i, j, d$ 为参数集 E 中参数个数^[13]. 得出 C_{ij} 为对象 o_i 在 C_{ij} 个属性上取值 $\geq o_j$, 通过计算对照表 1, 得出结果如表 2 所示.

表 1 系统软集信息表

U	e_1 $\omega_1=0.3$	e_2 $\omega_1=0.2$	e_3 $\omega_1=0.1$	e_4 $\omega_1=0.4$
o_1	0.5	0.7	0.3	0.9
o_1	0.8	0.4	0.2	0.5
o_3	0.6	0.8	0.6	0.7
o_4	0.3	0.4	0.4	0.6

表 2 软集合对照表

U	o_1	o_2	o_3	o_4
o_1	4	3	1	3
o_2	1	4	1	2
o_3	3	3	4	4
o_4	1	3	0	4

表 2 中第 i 行总和标记为 r_i , 计算公式如下:

$$r_i = \sum_{j=1}^n c_{ij}, i = 1, 2, \dots, n. \tag{1}$$

第 j 列总和标记为 t_j , 计算公式如下:

$$t_j = \sum_{i=1}^n c_{ij}, i = 1, 2, \dots, n. \tag{2}$$

每个属性 o_i 赋予一个权重 S_i , 计算公式如下:

$$S_i = r_i - t_i, i = 1, 2, \dots, n \quad (3)$$

4.4 计算各属性权重的方法

应聘者对用人单位的要求有一定的侧重性选择, 同时用人单位对应聘者的素质要求也有侧重, 因此需要针对不同的对象设定不同的权重. 本文以用人单位已经确定的各要素权重为例进行说明, 本信息系统以工作岗位为指标去区分每个要素的权重, 对同一岗位的不同要素通过成对相比法得到各个属性的权重值. 经计算得到的结果如表 3 所示.

表 3 各工作岗位权重表

相关因素	性别 X_1	年龄 X_2	学历 X_3	技术职称 X_4	...	$\sum x_n$
工作岗位						
技术工	0.6	0.4	0	0	...	1
办公室人员	0.2	0.3	0.3	0.2	...	1
车工	0.5	0.3	0.1	0	...	1
项目经理	0.1	0.1	0.4	0.4	...	1
...						1

各工作岗位的权重值计算步骤如下:

(1) 统计用人单位各指标的关注度

信息系统通过在数据库中统计曾成功就业的应聘者信息、用人单位招聘信息, 根据每个指标被用人单位关注的次数占所有招聘次数的比例去确定该指标的关注度, 具体公式如下.

$$\text{指标关注度 } \alpha_m = \frac{\sum \text{用人单位选择有要求次数}}{\text{总次数}} \quad (4)$$

(2) 使用成对相比法确定权重, 如表 4 所示

表 4 各要素成对比较法计算表

	性别 X_1	年龄 X_2	学历水平 X_3	...	元素 X_n
性别 X_1	1	a_{x1}, a_{x2}	a_{x1}, a_{x3}	...	a_{x1}, a_{xm}
年龄 X_2	a_{x2}, a_{x1}	1	a_{x2}, a_{x3}	...	a_{x2}, a_{xm}
学历 X_3	a_{x3}, a_{x1}	a_{x3}, a_{x2}	1	...	a_{x3}, a_{xm}
技术职称 X_4	a_{x4}, a_{x1}	a_{x4}, a_{x2}	a_{x4}, a_{x3}	...	a_{x4}, a_{xm}
...				...	
元素 X_n	a_{xn}, a_{x1}	a_{xn}, a_{x2}	a_{xn}, a_{x3}	...	1
合计	W_{X_1}	W_{X_2}	W_{X_3}		W_{X_N}

$$\omega_{x1} = 1 + \frac{a_{x2}}{a_{x1}} + \frac{a_{x3}}{a_{x1}} + \frac{a_{x4}}{a_{x1}} + \dots + \frac{a_{xm}}{a_{x1}}, \quad (5)$$

$$\omega_{x2} = \frac{a_{x1}}{a_{x2}} + 1 + \frac{a_{x3}}{a_{x2}} + \frac{a_{x4}}{a_{x2}} + \dots + \frac{a_{xm}}{a_{x2}}, \quad (6)$$

$$\omega_{xm} = \frac{a_{x1}}{a_{xm}} + \frac{a_{x2}}{a_{xm}} + \frac{a_{x3}}{a_{xm}} + \frac{a_{x4}}{a_{xm}} + \dots + 1. \quad (7)$$

通过以上公式可以得到各个工作岗位的指标权重向量 $\omega = (\omega_{x1}, \omega_{x2}, \dots, \omega_{xm})$

随着信息系统中数据库成功推荐的数据增多, 系统自动更新权重向量从而达到 $\omega = (\omega_{x1}, \omega_{x2}, \dots, \omega_{xm})$ 的取值更加准确和合理.

4.5 优化的软集理论推荐算法

在用人单位招聘和应聘者求职过程中引入软集的推荐算法, 可以极大提高成功率. 这两种算法原理一致. 本文以向用人单位推荐人才的“基于软集理论人才推荐算法”进行说明.

录入: 应聘者信息 $p(p_1, p_2, \dots, p_n)$, 用人单位 o , 工作岗位权重向量 $\omega = (\omega_{x1}, \omega_{x2}, \dots, \omega_{xm})$

输出: 推荐的用人单位列表及排名 L

(1) 由 4.2 节方法把应聘者信息 $p(p_1, p_2, \dots, p_n)$ 量化为软集信息表;

(2) 利用正则化参数约简法进行约简;

- (3) 由得到的软集图表根据 4.3 节公式得出软集合对照表;
- (4) 由公式(1)到(3)得出步骤(3)中对照表的 $S_i, i = 1, 2, \dots, c$;
- (5) 对 $S_i, i = 1, 2, \dots, C$ 进行排序得出推荐表 L ;
- (6) 把推荐表 L 中前 20 个值作为个人应聘者推荐给用人单位^[14].

5 人才推荐系统的实现

本章节将从两个方面来实现人才推荐系统,首先对系统的使用进行描述,然后对系统中的算法进行分析,下面分别说明.

5.1 系统使用过程

在系统开发的过程中,主要是用 Java 语言开发的,数据库则采用当今最流行的 oracle 数据库,因为系统是面向整个市场,数据量是非常庞大的.登入系统之后,呈现给求职人员的界面是一个表单,这个表单是给求职人员提供一个方便快捷的信息查找模式,通过对表单中求职基本信息的填写,可以使得系统查找更加精确,也使得企业可以更好地找寻自己需要的人才.

当求职人员填写完这些信息之后,系统会将这些信息上传到服务器,后台服务器会通过一些关键字在数据库中进行检索,找到符合的企业信息,然后回传至服务器,服务器再将这些企业信息返回给求职人员,求职人员可以点击自己想查看的企业信息^[15].

5.2 系统算法分析

这一部分主要是对系统的性能进行分析的过程,首先对系统推荐算法进行测试,测试的数据源来自郑州市人力资源与社会保障中心.测试的过程是以随机取样的方式进行的,分别在 3 组数据中随机抽取 100 个求职人员和 100 家招聘企业,每组数据都是以求职人员数量和招聘企业数量为一组,分别是 2000 和 5000、10 000 和 30 000、100 000 和 130 000.然后分别计算系统对双方的响应时间,如表 5 所示.接下来对系统推荐成功率进行分析,这个分析的数据来源是系统在郑州市运行半年时间所得数据.而系统验证推荐成功的数据来源则是通过求职者和企业反馈、各个人才市场的数据统计得到的.当在 3 个月内,系统中求职人员和推荐企业中的一家形成劳动关系,那么就称为系统推荐成功.以下是系统推荐成功率数据表 6.

表 5 系统响应时间表

组合	2000/5000	10 000/30 000	100 000/300 000
向求职人员响应时间/s	0.38	0.56	1.59
向招聘企业响应时间/s	0.35	0.53	1.56

表 6 系统推荐成功率数据表

种类	求职	招聘
系统推荐	6152	2459
推荐成功	5108	2178
成功率	0.83	0.88

从这两个测试,可以看出系统性能从整体上来说还是不错的,当求职人数增加时,系统响应时间并没有延迟太久,所以系统在处理数据时,性能还是处于一个比较高的水平的;从系统推荐成功率数据表中,可以看出不管是系统向求职者推荐还是向企业推荐,成功率都是比较高的,这说明系统在运行过程中是比较理想的.

6 总 结

系统通过(某方法)对毕业生信息和企业招聘信息进行对比碰撞,从而根据碰撞出的相同关键字,分别向毕业生和企业互相推荐,从而造就了一个新的求职招聘模式.但是由于求职者和招聘企业在互相选择时的一些主观意愿的改变,造成了一些信息的废弃,从而使得系统推荐的准确度下降,这也是系统接下来需要改进的地方.从系统目前的测试情况来看,还是比较符合目前的毕业生求职和企业招聘的需求的.

参 考 文 献

- [1] 王 乐. 高校就业现状与就业资源整合研究[J]. 湖北函授大学学报, 2015(8): 3-4.
- [2] 李木昌. 基于 B/S 模式毕业生就业系统的设计与实现[J]. 软件导刊, 2014, 12: 89-91.

- [3] 耿生玲,李永明,刘震.不完备决策软集与优势可信规则获取[J].计算机工程与科学,2013(12):153-160.
- [4] 毛军军,姚登宝,王翠翠,陈华友.基于时序模糊软集的群决策方法[J].系统工程理论与实践,2014(1):182-189.
- [5] 陈文,余本功.基于Vague软集的模糊群决策方法研究[J].计算机工程与应用,2014(7):104-107.
- [6] 邵亚斌,张晓霞,徐永琳.基于模糊软集和加权模糊软集的群决策方法[J].统计与决策,2014,19:46-49.
- [7] 刘庆,王昌.基于Vague软集相似度量的快速估算模型[J].河北大学学报:自然科学版,2014(5):460-466.
- [8] 耿生玲,李永明,刘震.关联规则挖掘的软集包含度方法[J].电子学报,2013(4):804-809.
- [9] 耿生玲,李永明,冯峰.软集决策信息系统的属性约简[J].小型微型计算机系统,2011(4):721-725.
- [10] 郭智莲,杨海龙.区间值模糊软集的格结构[J].计算机工程与应用,2011(33):7-9.
- [11] 滕艳辉,王昌.多目标模糊决策的Vague软集法[J].计算机工程与应用,2012(10):6-8.
- [12] 尤天慧,李铭洋,樊治平,曹兵兵.基于软集理论的考虑评价指标集有差异的政府采购评标方法[J].系统工程,2012(12):105-109.
- [13] 计艳飞,李生刚.对偶软集和对偶软子坡代数[J].计算机工程与应用,2014(12):1-6.
- [14] 陈秀明,钱丽,李敬明,等.广义区间梯形模糊软集在群偏好集结中的应用[J].计算机应用,2014(12):3451-3457.
- [15] 宋旭东,李佳骏,李艳红.Vague软集的分解定理、表现定理和扩张原理[J].计算机工程与应用,2015(6):99-102.

Research of Graduates Recommendation Platform Driven by Soft Set Theory

YANG Jiayi

(College of Information Science, Zhengzhou Institute of Aeronautical Industry Management, Zhengzhou 450015, China)

Abstract: Although network technology development serves as an effective link between graduates and enterprises, whereas most recruitments and recommendation platforms are still designed to perform respectively, lacking of effective joints. Catering to the direct needs and core requirements of graduates and enterprises, an intelligent and decision-supporting graduates recommendation platform is constructed by soft set theory. Recommendation platform allows users to set self-care properties and rank properties, therefore, recommendation results can be sorted in accordance with quality to make them become more objective and comprehensive as well as meeting the needs of college students and enterprises. Finally, experimental results show that the method contains high performance.

Keywords: intelligent recommendation; soft set; decision-making; graduates recommendation; platform