

大连海事大学

硕士学位论文

网络招聘信息个性化推荐技术研究

姓名：张维瑞

申请学位级别：硕士

专业：计算机科学与技术（计算机软件与理论）

指导教师：孟丽艳;林正奎

20100601

## 摘 要

1993 年伴随着美国政府 National Information Infrastructure (简称 NII) 计划的实施, 运用互联网及相关技术向招聘者和求职者提供信息服务的新型人力资源中介服务方式, 即网络求职与招聘应运而生。如今, 随着互联网技术与应用的日益普及, 加之网络求职招聘在信息传播速度、服务成本、时间投入、便捷性等方面较之传统方式所具有的自身优势, 网络求职招聘得到了人们更多的青睐, 并已成为最重要的求职招聘方式之一。

与此同时, 近年来随着网络招聘的日益普及, 参与者规模迅速扩大, 信息爆炸式增长以及由此而引发的信息迷航问题日益严重, 加之就业竞争压力的日益加剧和招聘实效性的日益增强, 求职者已不在满足于简单的招聘信息和求职简历的网络化分享, 进而对网络求职招聘在个性化服务方面提出了更高的要求。

本文以 Soelberg 和 Kanfer 关于求职过程的定义为理论依据, 即将求职行为看成一种指向目标的活动而非单纯的职位信息搜索, 目的是为了缩小求职者目前状态与就业目标之间的差距。进而站在求职者的角度, 围绕网络招聘信息的个性化推荐技术进行了相关探索性的研究工作, 以期帮助求职者以更准确、更即时的方式获取职位信息, 并采取更加理性的方式进行个人求职简历的投递。主要研究内容包括: 网络招聘求职者求职过程机理模型研究、网络求职招聘信息个性化推荐策略研究, 网络招聘求职者求职意愿模型研究、网络招聘求职者职位适合度评价模型研究以及网络求职者之间职位竞争力评价模型研究和网络招聘信息的个性化推荐技术研究。在此基础上, 通过网络招聘信息个性化推荐原型系统的设计与实现, 对论文的研究成果进行了初步验证。

个性化推荐技术是近年来刚刚兴起的信息服务技术, 受到情境复杂性的影响, 其研究具有一定的挑战性。由于时间及自身水平有限, 论文研究工作尚缺乏深度, 且存在很多不足之处, 作者将在后续的研究过程中加以持续改进。

**关键词:** 网络求职招聘; 个性化服务; 多属性决策

## ABSTRACT

In 1993, along with the United States Government National Information Infrastructure (referred to as NII) plan, the use of Internet and related technologies to offer recruiters and job seekers a new human resources information services intermediary service mode, that is, jobs and recruitment network came into being. Now, with the application of Internet technology becoming increasingly popular, coupled with advantages of job recruitment network in information dissemination speed, service costs, time investment, the network job recruit is favored by more and more people, and has become the most important recruitment methods.

At the same time, in recent

years with the growing popularity of online recruitment, participants expanded rapidly and the explosive growth of information and the resulting information confusion lead to a series of problems. With increasing competitive pressures of employment and recruitment, current job seekers are no longer satisfied with simple job search methods like sending out resumes and the resume network share, individuals have higher expectation to the network job recruiting services.

Based on Soelberg and Kanfer's theoretical basis on the job search process definition, which is, the upcoming job search behavior is a goal-oriented activities rather than simply the job information search, the purpose is to reduce the gap between the status quo and the job expectation. Standing in the applicant's point of view, this paper explored relevant technique about online recruitment information, personalized recommendation, aiming to help job seekers getting information with high accuracy and more immediate, and to adopt a more rational approach to personal resume delivery. Several works are carried out in this research: network job seekers' recruitment process model, strategies for network job personalized recommendation, network job seekers' will model, evaluation model of job seekers' recruitment results, network job seekers' competitiveness evaluation model and online recruitment personalized recommendation techniques. On this basis, a personalized recommendation prototype system is designed and implemented to evaluate the aforementioned work.

Personalized recommendation technology is just emerging in recent years. Influenced by complex situations, personalized recommendation is quite a challenging work. Due

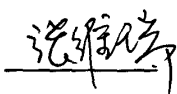
to time and personal limitation, the research may have inadequacies, which will be improved in the follow-up study.

**Key Words: Network Recruitment ; Personalized recommendation; Multiple Attribute Decision**

# 大连海事大学学位论文原创性声明和使用授权说明

## 原创性声明

本人郑重声明：本论文是在导师的指导下，独立进行研究工作所取得的成果，撰写成博/硕士学位论文 “网络招聘信息个性化推荐技术研究”。除论文中已经注明引用的内容外，对论文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本论文中不包含任何未加明确注明的其他个人或集体已经公开发表或未公开发表的成果。本声明的法律责任由本人承担。

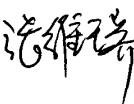

学位论文作者签名： 

## 学位论文版权使用授权书

本学位论文作者及指导教师完全了解大连海事大学有关保留、使用研究生学位论文的规定，即：大连海事大学有权保留并向国家有关部门或机构送交学位论文的复印件和电子版，允许论文被查阅和借阅。本人授权大连海事大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，也可采用影印、缩印或扫描等复制手段保存和汇编学位论文。同意将本学位论文收录到《中国优秀博硕士学位论文全文数据库》（中国学术期刊（光盘版）电子杂志社）、《中国学位论文全文数据库》（中国科学技术信息研究所）等数据库中，并以电子出版物形式出版发行和提供信息服务。保密的论文在解密后遵守此规定。

本学位论文属于： 保 密 ☐ 在 \_\_\_\_\_ 年解密后适用本授权书。

不保密 ☐ （请在以上方框内打“√”）

论文作者签名：  导师签名： 

日期：        年    月    日

## 第1章 绪论

### 1.1 研究背景及意义

所谓招聘即根据企业、机构、组织的运营框架、岗位设置需要等面向组织外部征集应聘者以获取人力资源的过程，是组织根据自身发展需要，向外界发布招聘信息，并对应聘者进行有关测试、考核、评定及一定时期的试用，综合考虑其各方面条件之后决定企业的聘用对象的常用方式。

求职则是与招聘相对应的过程，从职业探索的角度来看，目前关于求职行为的主要定义可以概括为以下四种：

定义 1：Steffy、Shaw 和 Noe 将求职行为定义为“对职业信息的数量和适用性进行探索的行为，基本上可以看作是一种信息搜寻过程”<sup>[1]</sup>。

定义 2：Bretz、Boudreau 和 Judge 将求职行为定义为“花时间和精力获取有关劳动力市场信息的特定行为，无须考虑动机因素及求职结果，只是一种信息搜集活动”<sup>[2]</sup>。

定义 3：Soelberg 认为，求职是一系列彼此相关的四种行为：识别理想职业、制订搜寻计划、搜寻并选择工作、确定并接受选择结果<sup>[3]</sup>。

定义 4：Kanfer 认为，求职行为是一种“动态的、循环的自我调适过程，是一种有目的并受主观意愿驱动的行为模式，它开始于就业目标的识别，进而为实现就业目标付出努力”<sup>[4]</sup>。

上述四种关于求职行为定义的共同点即都将职位信息搜寻看做求职过程的基础和重要环节。在此基础上，定义 3 和定义 4 从行为适应领域对求职行为定义进行了扩展，将求职行为看成一种指向目标的活动，目的是为了缩小目前状态与就业目标之间的差距，即认为求职行为不仅包括搜寻信息，而且包括为申请职位而制订决策、确定并接受工作等一系列行为，是一种以解决问题为导向的应对策略。在实践过程中，定义 3 和定义 4 更加符合现代求职者，尤其是高端求职者的行为特征。

鉴于求职过程中职位信息的重要作用，围绕职位信息的发布与搜索，求职招聘经历了从以招聘会、报纸、电视等为媒介的传统求职招聘到以互联网为媒介的网络求职招聘的方式转变。尤其近年来，随着互联网的快速普及，加之网络求职

招聘在信息传播速度、招聘成本、时间投入、便捷性和招聘成功率等方面较之传统招聘方式所具有的自身优势，网络求职招聘得到了人们更多的青睐，2004年，国内网络招聘占整个招聘市场容量的13.2%，2005年网络招聘市场成长迅猛，占招聘市场份额接近20%，2006年这一比例达到了约28.2%，2008年这一比例达到了约54.2%<sup>[5]</sup>，表明网络招聘已经发展成为最主要的招聘方式之一。

所谓网络招聘通常指运用互联网及相关技术，向招聘者和求职者提供招聘过程服务，包括招聘信息发布、简历下载、定制招聘专区、求职简历生成、职位搜索等。近年来，随着网络招聘日益普及，参与者规模迅速扩大，人们对网络招聘提出了更高的要求，主要表现在以下几个方面：

### （1）高效搜索

由于互联网的飞速发展，网络求职招聘的用户数和资源数都呈现出巨量性和海量性，使用关键词进行搜索后，动辄得咎上千个结果，如果输入关键词组合后，得到的结果会更多，浪费用户的时间资源和网络资源。人们对高效搜索信息提出了更高的要求。

### （2）招聘信息即时获取

目前的网络求职招聘一般要求求职者在线状态下，进行搜索，才能得到信息，而如果求职者处于离线状态，则不能在第一时间获取发布的新信息，这可能会错失工作机会。

### （3）准确投递简历

现在的网络求职招聘面临的普遍问题是，用户按自己需求搜索到的信息量特别大，几乎无从选择，因此迫切希望有一种辅助决策机制来帮助其进行职位选择。

导致上述需求产生的原因是，目前人们对网络求职招聘行为的认识还停留在定义1和定义2阶段，而忽视求职过程的本质内涵。有鉴于此，近年来，部分学者提出将“个性化信息服务”<sup>[6]</sup>引入到网络求职招聘中，以期使求职者更准确获取信息，从而能够准确投递简历。

基于以上背景，研究“网络招聘信息个性化推荐技术”对网络求职招聘具有重要的现实意义，同时对个性化信息服务也具有重要的理论研究价值。

## 1.2 研究内容与技术路线

### 1.2.1 研究内容

基于对求职定义 3 和定义 4 的认同，从求职者求职行为过程机理分析入手，开展网络招聘信息个性化推荐技术研究，主要包括：

- (1) 求职者求职过程模型研究；
- (2) 求职者职位竞争力评价模型研究；
- (3) 求职者职位适合度评价模型研究；
- (4) 求职者求职意愿模型研究；
- (5) 网络招聘信息个性化推荐原型系统的设计与实现。

### 1.2.2 技术路线

#### (1) 文献收集与整理

紧密跟踪国内外关于个性化信息推荐技术、网络求职招聘技术研究的最新成果，进行文献收集与整理工作，为后续研究提供理论基础。

#### (2) 关键技术研究

围绕“求职者求职过程模型研究”、“求职者职位竞争力评价模型研究”、“求职者职位适合度评价模型研究”、“求职者求职意愿模型研究”开展探索性研究工作。

#### (3) 研究成果验证

设计、开发网络招聘信息个性化推荐原型系统，对论文研究成果进行初步验证。

技术路线图如图 1.1 所示：



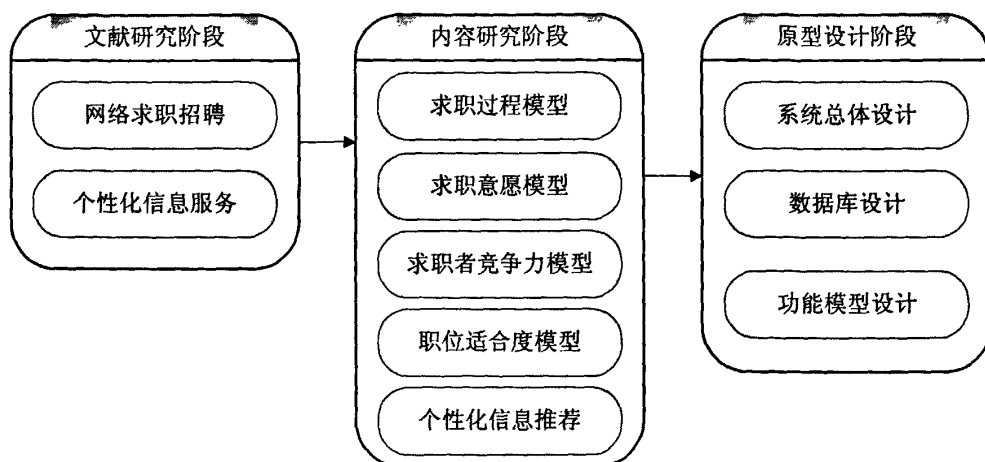


图 1.1 技术路线图

Fig 1.1 technic route chart

### 1.3 论文结构

本论文共分为 5 章，具体组织方式如下：

第一章 绪论，简要介绍了课题的研究背景及意义、论文的主要研究内容、技术路线及论文的组织结构。

第二章 相关问题研究综述，介绍了网络求职招聘和个性化信息推荐服务的国内外研究现状，并对个性化服务中服务对象信息需求判定、信息资源整合和基于服务对象信息需求与信息资源匹配的个性化信息推荐三个基本环节进行了详细的介绍。

第三章 招聘信息个性化推荐策略设计与关键技术，介绍了“求职者求职意愿模型”、“求职者求职过程模型”、“求职者职位竞争力评价模型”、“求职者职位适合度评价模型”的定义、作用与构建过程。

第四章 原型系统设计与实现，主要介绍了网络招聘信息个性化推荐原型系统的工作原理、总体设计、数据库设计及各个模块的设计与实现，最后通过一个实例进行验证。

第五章 结论与展望，对本文的工作进行了总结，并对下一步的研究工作提出一些设想。

#### 1.4 小结

本章主要介绍了课题的研究背景及意义、论文的研究内容、技术路线及论文的组织结构。

第 2 章 相关问题研究综述

2.1 网络求职招聘

2.1.1 国内外网络求职招聘的产生及发展

(1) 国外网络求职招聘的产生及发展

1993 年美国政府实施 National Information Infrastructure（简称 NII 计划），随之产生了利用互联网为招聘者和求职者提供信息服务的方式，与此同时网络用户数量迅速增加，专业求职招聘网站也应运而生<sup>[7]</sup>。在网络求职招聘网站出现初期，简历只能通过传真接收，而且没有自动或批量的职位发布方式。1997 年是人才网站大量兴起的时期，以 HotJobs 为代表的人才网站虽然只面对专业人才提供职位信息，但是已经实现了在线招聘。到 2005 年以 Careerbuilder 为代表的专业性、综合性的大型人才招聘网站越来越成熟；2006 年小型行业型招聘网站兴起，对大型网站造成很大压力，二者的竞争也越来越激烈。现在网络招聘已经普及到各个行业，并且求职者数量也越来越多<sup>[7]</sup>。

表 2.1 国外网络求职招聘的发展历程

Tab.2.1 The Development of overseas network job recruitment

时间（年）	关键词	典型代表	求职用户	特点
1996	起步	Monster	300 万	简历只接收 传真格式
1997	大量兴起	HotJobs	专业人才	在线招聘
2005	大型网站	Careerbuilder	综合人才	综合性职位，满足各 种岗位需求
2006	行业型网站	Linkedin	各类人才	在线招聘人数超过 传统招聘
目前	网络招聘与传 统招聘相结合	Careerbuilder	普及	各种招聘网站并存

(2) 国内网络求职招聘的产生及发展

国内与国外的网络求职招聘几乎同时起步，1997 年中国出现了以“中华英才网”为代表的第一批网络求职招聘网站<sup>[7]</sup>，经过十余年的发展，网络求职招聘已经

成为人们求职招聘的主要途径之一。

表 2.2 国内网络求职招聘的发展历程

Tab.2.2 The development of domestic network job recruitment

时间（年）	关键词	典型代表	求职用户	特点
1997	兴起	中华英才网	较少	多种招聘形式共存
1999	迅速增长	前程无忧网	增长	人才招聘网站多
2003	全国大型网站	/	继续增长	网上求职火爆
2005	地方性综合网站	南方人才网	人数骤增	专业化程度提高
2007	行业类网站	CSDN	有所下降	各种网站并存
目前	应届生求职网	我是应届生	学生多	竞争激烈

2.1.2 网络求职招聘的优点及特点

(1) 优点

网络求职招聘的优势主要表现在以下几个方面<sup>[5]</sup>：

- 1) 求职者与招聘者之间信息对称；
- 2) 方便、快捷、时效性强；
- 3) 信息量大；
- 4) 经济实惠；
- 5) 无地域限制；
- 6) 提高工作效率，缩短招聘周期。

因此近年来，网络求职招聘成为人们求职招聘的主要途径之一。

(2) 特点

通过对国内外代表性网络招聘平台的研究分析，现阶段网络招聘具有如下特点：

- 1) 在理念上，基于求职的定义 1 和定义 2，将求职行为认定为一种花时间和精力进行信息搜寻的过程。
- 2) 在技术上，依托于网络招聘平台开展网络服务。
- 3) 在服务内容上，呈现出同质化趋势，普遍包含招聘信息发布、简历下载、定制招聘专区、求职简历生成、职位搜索等模块。
- 4) 在客户体验上

根据 2010 年 1 月中国互联网发展状况统计报告 (CNNIC)<sup>[8]</sup>显示, 20~30 岁的高校学生占据中国网民的很大一部分。据统计<sup>[7]</sup>, 2003 年我国高校毕业生人数为 376 万, 2004 年达到 509 万, 2006 年已经超过 900 万人, 并且他们也是网络求职招聘的主要参与者。从学生的角度看, 海量信息引发的信息迷航使求职者不能高效搜索、招聘信息不能即时获取、并且没有相应的辅助决策机制帮助求职者准确投递简历。

## 2.2 个性化信息服务

个性化信息服务<sup>[6]</sup>是一种网络信息服务的方式, 旨在根据用户的需求, 借助于计算机及网络技术, 对信息资源进行收集、整理和分类, 向用户提供和推荐相关信息, 以满足用户对信息的个性化需求。

个性化信息服务包括服务对象信息需求判定、信息资源整合和基于服务对象信息需求与信息资源匹配的个性化信息推荐三个基本环节, 即为了实现个性化服务, 先要跟踪和学习用户的兴趣和行为, 并设计一种合适的表达方式, 同时为了把资源推荐给用户, 也必须组织好相应的资源, 并采用合适的信息匹配算法, 向用户推荐个性化信息。

### 2.2.1 服务对象信息需求判定

服务对象信息需求判定主要用来描述用户的兴趣和行为<sup>[7]</sup>, 它是个性化服务的基础, 主要包括用户需求、数据收集方式、数据组织方式及信息更新方式<sup>[9]</sup>。

#### (1) 用户需求

根据应用背景不同, 用户需求信息也各不相同<sup>[8]</sup>, 比如 smartpush<sup>[10]</sup>的应用领域是报纸, 用户可能关心的报纸类别是体育、娱乐、新闻等, FireFly<sup>[11, 12]</sup>的应用领域是音乐, 用户可能关心音乐风格, 歌手等, Amazon.com<sup>[13, 14]</sup>推荐的是书籍, 用户可能关心书作者、书类别、书内容等, 还有一些不是面向特定领域的, 只是用来推荐、帮助和搜索等。不管是哪种应用背景, 首先都要深入了解该领域的背景才能很好的描述用户的需求。

根据信息的变化状态, 将其分为静态数据和动态数据。静态数据包括用户的姓名、性别、年龄、教育背景和兴趣等在一定时间内不变化的信息; 动态信息包括用户浏览资源时产生的动作, 如浏览页面点击次数、页面停留时间、收藏、保

存等实际行为<sup>[15,16]</sup>。

确定了用户需求属性后，就要考虑这些数据是否可获得。下面一小节就来介绍数据的收集方式。

## （2）数据收集方式

数据收集是一个获取与用户特征、偏好或活动相关信息的过程<sup>[9]</sup>。根据数据是否需要用户直接参与，将数据收集方式分为两种，分别是显性收集方式和隐性收集方式。

显性收集方式是指用户根据系统要求，主动输入相关信息，从而达到学习的目的<sup>[17]</sup>。这种方式是比较直接而简单的做法，一般在用户第一次使用系统时，系统大多都会要求用户先注册自己的基本信息和感兴趣的内容，另外在使用过程中，系统可以要求用户反馈自己对推荐资源的喜好程度等。但一般情况下这种方法只能收集一些简单的信息，很难收到实效。因为用户怕个人信息泄露、过多输入信息麻烦等原因放弃登录该网站也是可能的。所以大多数网站比较偏向于信息的隐性收集方式。

隐性收集方式是指不需要用户直接输入什么信息，而是通过跟踪用户行为，比如查询、标记书签、点击鼠标、复制、收藏等不同的行为可揭示用户不同的兴趣<sup>[18]</sup>。如文献 18 的研究表明，用户浏览页面所花的时间不同，可以表明用户对其感兴趣程度不同，标记书签、收藏夹中的信息可以有效的提示用户的兴趣。

随着数据挖掘技术的不断发展，通过分析 Web 日志来获得用户的兴趣，也是隐性收集方式的一种重要方法。日志信息包括用户 ID、IP 地址、客户端访问方式等信息，比如 Ligson、web Personalizer<sup>[19]</sup>是通过这类信息的挖掘来发现用户感兴趣的知识。

隐性收集方式在不增加用户负担的情况下获取数据，因此比显性收集方式更加可行。但是一般情况下隐性数据中可能包含无关信息，增加系统复杂度。

由于单一方式各有优缺点，因此一般采用二者结合的混合收集方式。

## （3）数据组织方式

收集到的数据如何表示，目前并没有统一的标准，常用的表示方法有以下几种：

### 1) 基于向量空间模型的表示<sup>[20]</sup>

基于向量空间模型是目前比较流行的用户模型表示方法。如 Personal WebWatcher<sup>[21]</sup>、Syskill&Webert<sup>[22]</sup>、Letizia<sup>[16]</sup>等都采用这种向量空间模型表示方法。在该模型中,每一篇文档都被映射成多维向量空间中的一个点,对于所有的文档类和未知文档,都可用此空间中的向量( $T_1, W_1; T_2, W_2; \dots; T_m, W_m$ )来表示(其中  $T_i$  为词,  $W_i$  为词对应的权值,用以刻画该词在描述此文档内容时的重要程度),从而将文档信息的表示和匹配问题转化为向量空间中向量的表示和匹配问题来处理。

## 2)基于本体论的表示<sup>[20, 23]</sup>

本体论(Ontology)也称为“本根论”,是一个哲学的概念,用来描述事物的本质。Studer 等人对本体进行了深入的研究,认为本体是共享概念模型的明确的形式化规范说明<sup>[24]</sup>。本体的目标是获取、描述和表示相关领域的知识,提供对该领域知识的共同理解,因此可以用本体来描述用户的兴趣。比如文献<sup>[25]</sup>开发设计了一个基于本体的个性化智能信息检索系统模型。本体通常采用概念树的形式,层次之间的关系有 part-of、kind-of、instance-of、attribute-of<sup>[24]</sup>等。

此方法最大的好处是能够实现知识的重用和共享,但本体的设计主要依赖专家知识,因此主观性比较强。

## 3)基于用户-项目评价矩阵的表示<sup>[26]</sup>

该模型用一个  $T_m \times n$  矩阵来表示用户模型,其中  $m$  代表用户数量,  $n$  代表资源数目,矩阵元素  $T_{ij}$  表示用户  $i$  对资源  $j$  的评价,如果采用这种表示方法,后续的推荐算法多采用协同过滤算法,如:Recommender、Video、Ringo 和 GroupLens 等都采用基于用户-项目评价矩阵来表示用户模型。

基于用户-项目评价矩阵的表示方法比较简单、直观,但是随着用户数量和资源数量的不断增加,矩阵稀疏性问题会变得越来越严重,后续计算也比较困难。

## 4)XML 文件描述<sup>[9, 17]</sup>

XML (Extensible Markup Language) 即可扩展标记语言,它与 HTML 一样,都是 SGML(Standard Generalized Markup Language,标准通用标记语言)。XML 是 Internet 环境中跨平台的,依赖于内容的技术,是当前处理结构化文档信息的有力工具。

XML 与其它数据库不同, 它提供了更强有力的数据存储和分析能力, 例如: 数据索引、排序、查找、相关一致性等。并且 XML 的扩展性强, 可以创建个性化的标记语言, 称之为元语言。XML 的标记语言可以自定义, 这样可以提供更多的数据操作, 而不像 HTML 一样, 只能局限于按一定的格式在终端显示出来。

通过以上分析可知, 用户模型的各种数据组织方法各有利弊, 我们应该根据实际应用背景选择合理的表示方法。另外, 研究独立于模型的通用的用户模型表示方法也是用户模型表示中的一个重要研究内容。

#### (4) 信息更新方式

根据用户对推荐资源的显性或隐性的反馈来更新用户模型, 确保用户模型信息能够随着用户的喜好变化而变化, 提高用户模型的精度和推荐质量, 也是研究的重要内容之一。

目前采用的更新技术主要分为三类, 分别是神经网络技术、遗传算法和信息增补技术<sup>[9]</sup>。

##### 1)神经网络技术

其原理是通过调整网络连接权重来自适应更新用户描述文件。当用户的偏好随着时间的变化发生变化时, 神经网络自适应的调整网络连接权重, 更新网络输出。一类是不增加类别, 而在原有类别的基础上调整分类权重, 另一类是增加新的类, 并删除过时的知识类来应对用户新兴趣的兴起<sup>[27]</sup>。

##### 2)遗传算法

遗传算法 (Genetic Algorithm)<sup>[28]</sup>是模拟达尔文生物进化论的自然选择和遗传学机理的生物进化过程的计算模型, 是一种通过模拟自然进化过程搜索最优解的方法, 它最初由美国 Michigan 大学 J.Holland 教授在 1975 年首先提出来的。

该算法由染色体种群、适应度函数和选择、交叉、变异三个主要操作算子组成。每一代种群包含了若干个个体, 被称为染色体。根据每个个体的适应度函数值, 经过选择、交叉和变异操作得到新的种群, 从而使模型得到更新。

##### 3)信息增补技术

信息增补技术主要是从用户反馈中抽取新信息添加到用户模型中。这是目前为止使用比较多的一类用户模型更新技术。这种技术又包括涉及权重调整的信息



增补和直接的信息增补。

前者不仅增加新的关键词，而且还会调整用户模型中新、旧信息的权重，使新信息在推荐中起到作用。同时无效的旧信息随着权重的减小最终从模型中删除。如 Personal Web Watch<sup>[29]</sup>、WebCobra<sup>[30]</sup>、Letiziat<sup>[16]</sup>等系统采用这种方式来更新模型。

后者将用户的反馈信息直接添加到用户模型中，典型系统有 GroupLens<sup>[31]</sup>、Websell<sup>[32]</sup>、Video Recommender<sup>[33]</sup>等。这种更新方法比较简单，但是只给用户模型增加新信息，而不删除旧信息，因此规模会不断增大，模型维护会越来越难。

总之，模型更新技术也受模型表示方法和后续推荐算法的限制，因此，应该根据具体应用来选择合理的更新技术。

### 2.2.2 信息资源整合

个性化服务所应用的领域决定了它所处理的资源。GroupLens 应用的领域是新闻；Amazon.com<sup>[13, 14]</sup> 应用的领域是电子商务，主要是推荐书籍；MovieLens<sup>[34]</sup> 应用领域是电影，还有一些是处理 Web 访问日志等，而到目前为止不管是推荐哪种资源，大都是将其转化为文本，比如 MovieLens 推荐电影，主要是通过电影的导演、主要演员、电影类别、主要内容等特征来描述电影，所以仍然属于文本处理。

资源描述文件可以用基于分类的方法和基于内容的方法来表示，下面分别介绍这两种方法。

#### (1) 基于分类的方法

基于分类的方法是通过将资源进行分类，然后将用户感兴趣的同一类文档推荐给用户。目前常用的文本分类的方法一般有三种<sup>[6]</sup>，分别是K最近邻方法(KNN)、朴素贝叶斯(Naïve-Bayes)以及支持向量机方法(SVM)。

#### (2) 基于内容的方法

基于内容的方法是从资源本身出发，提取能表示资源内容的信息来表示资源，其本质是进行文本特征提取，目前常用的方法主要有以下几种方式：1) 信息增益(information gain, IG)<sup>[9]</sup> 通过统计某个特征在一篇文档中出现或者不出现的次数来描述文档；2) 互信息(mutual information)<sup>[35]</sup>是计算每个特征和文档的相关

性；3) X2 统计方法，单松巍等人<sup>[34]</sup>对这几种方法进行了对比研究，结果显示第一种和第三种比较好，但是算法复杂度也比较大。

文档特征提取以后，就得计算每个特征的权重，最常用的方法是 TFIDF (Term frequency inverse document frequency)，该方法简单易行，且目前国内外很多学者对此方法进行了深入的研究，其效果已经得到很大提高。

### 2.2.3 个性化信息推荐

目前个性化推荐技术主要有两种，分别是基于规则的技术和信息过滤技术，其中信息过滤技术包括协同过滤技术和基于内容过滤的技术<sup>[6]</sup>。

#### (1) 基于规则的技术

基于规则的技术，主要是推荐给用户一些采用关联规则发现算法计算出的内容。关联规则的发现算法有很多，如 Apriori、AprioriTid、DHP、FP-tree 等。其优点是简单，易于理解，但缺乏通用性，而且随着规则的不断增加，系统会越来越难维护，因此目前实际采用这种方法的系统比较少。

#### (2) 信息过滤技术

信息过滤技术又可以分为协同过滤技术和基于内容过滤的技术。

协同过滤 (Collaborative Filtering, 简称 CF) 技术是最早被提出来并得到广泛应用的个性化推荐技术，它是根据用户之间的相似性来产生推荐结果。国内外很多学者对此算法进行了研究，如 Herlocker J 等人<sup>[36]</sup>提出了一种研究协同过滤算法的框架；Sarwar B 等人<sup>[37]</sup>提出了一种基于资源的协同算法；王虹予等人<sup>[38]</sup>将两种传统的协同过滤算法决策树算法和贝叶斯网络算法结合使用；吴丽花等人<sup>[9]</sup>主要分析了在推荐算法中的用户建模方法；崔林等人<sup>[39]</sup>提出了一种基于资源语义知识协同过滤算法等等。

协同过滤技术的优点是可以推荐各种难以进行内容分析的产品，比如音乐、电影等，但是由于这种技术的用户建模技术一般是基于打分矩阵的，所以对于打分稀疏性问题目前还没有好的解决方法，而基于内容过滤技术 (content-based filtering) 则不存在这个问题，它是通过分析用户已经浏览 (或者购买) 过的内容，建立用户兴趣模型，然后比较用户与产品之间的相似性，直接向用户推荐与之最相似的产品。

在基于内容过滤的技术中，用户兴趣模型的建立和更新是最为关键的部分，王宏宇等人<sup>[40]</sup>通过引入支持向量机方法分析项目内容，构建用户模型；冀俊忠等人<sup>[41]</sup>采用贝叶斯网方法建立用户模型；Somlo<sup>[42]</sup>等人提出了利用自适应过滤技术更新用户兴趣模型等。并且基于内容的推荐技术大多数都将产品内容描述成关键词，但是基于内容的推荐算法在内容特征提取技术上存在困难，如对于电影的推荐还是将其转化为文本描述来进行推荐。

信息过滤技术在计算相似性时采用的方法主要有夹角余弦公式和 Pearson 相关性公式。不同的系统会采用不同的方法计算相似性，而且还会对公式加以改进，以使结果更加准确。

个性化推荐技术由于受到其应用背景的影响，其研究具有一定的挑战性。因此本文以个性化推荐服务理论为基础，结合网络求职招聘的实际，提出以求职意愿模型、求职者竞争力模型和求职者职位适合度模型为基础，进行职位信息的个性化推荐服务。

## 2.3 结论

随着互联网的发展，网络求职招聘已经取代招聘会、报纸等传统招聘方式，成为最主要的求职招聘方式之一。但是伴随着求职者和招聘者的规模的越来越大，用户对高效搜索、信息即时获取、准确投递简历的要求越来越高，因此将个性化信息服务引入到网络求职招聘中，对于有效解决目前网络求职招聘所面临的新挑战具有重要意义。

## 2.4 小结

本章对网络求职招聘和个性化信息服务进行了文献综述。

### 第 3 章 招聘信息个性化推荐策略及关键技术

#### 3.1 求职者求职过程模型

Soelberg 认为，求职是一系列彼此相关的四种行为：识别理想职业、制订搜寻计划、搜寻并选择工作、确定并接受选择结果<sup>[3]</sup>。Kanfer 认为，求职行为是一种“动态的、循环的自我调适过程，是一种有目的并受主观意愿驱动的行为模式，它开始于就业目标的识别，进而为实现就业目标付出努力”<sup>[4]</sup>。

由于目前的网络求职招聘都将求职看作是一种信息搜索，因此一般只包括识别理想职业、制定搜索计划并搜索两部分，但实际上 Soelberg 和 Kanfer 有关求职的定义更符合求职的本质。因此，本文以 Soelberg 和 Kanfer 定义为理论依据，将求职定义为一个不断自我调试的过程，如图 3.1 所示：

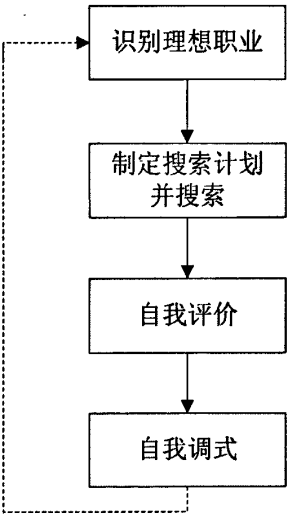


图 3.1 求职过程模型

Fig.3.1 The model of Job search process

为了识别理想职业，建立一个“求职者求职意愿模型”，为了给求职者提供“个性化信息”，建立“求职者职位竞争力评价模型”和“求职者职位适合度评价模型”。通过这两个模型，不仅能给求职者提供辅助决策机制，求职者还可以根据自己兴趣和爱好选择职位信息重组方式，并根据这两个指数进行自我评价，自我调式。如图 3.2 所示：

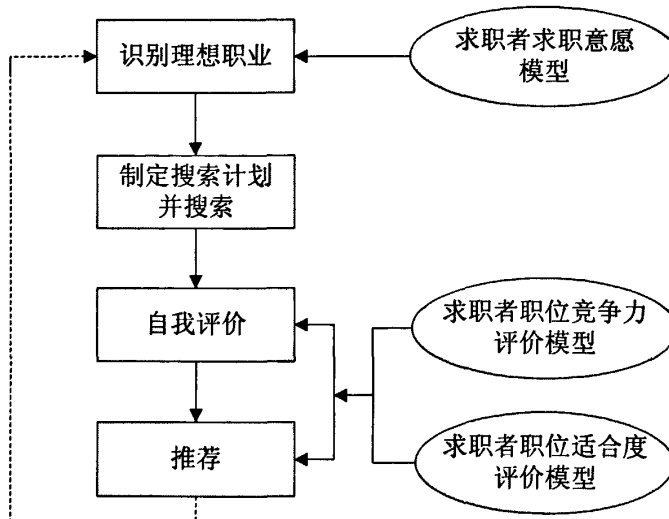


图 3.2 职位信息个性化推荐策略

Fig 3.2 Personalized recommendation strategy job information

## 3.2 求职者求职意愿模型

### 3.2.1 定义

求职者求职意愿模型用来综合反映求职目标，是一个动态的变化过程。

### 3.2.2 作用

主要作用是综合各方面信息匹配满足其要求的职位信息，所谓的“要求”也就是说求职者的求职意愿。同时，根据求职者的求职意愿对信息进行过滤，减少信息量，这样可以有效的解决网络招聘中信息迷航的问题。

### 3.2.3 模型构建方法

主要通过实证分析和聚类方法来构建求职者求职意愿模型。

本文运用实证分析——调查法对中华英才网、智联招聘网、前程无忧网等国内知名网络求职招聘网站的求职意愿属性进行调查，综合得到求职意愿属性的全集，如图 3.3 所示

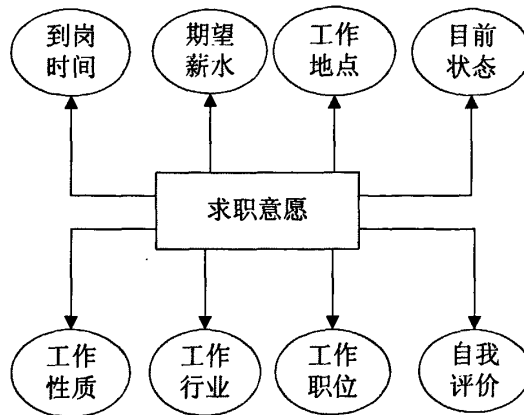


图 3.3 求职意愿属性

Fig 3.3 Job seeker intention attribute

经过聚类分析，“自我评价”主观性较强，且由于网络求职者人数众多，自我评价各不相同，侧重点也不一样，因此为简单起见，本文在此模型中不考虑自我评价这个因素。另一方面，本文所做的研究服务对象主要是应届毕业生求职者，因此“目前状态”一项均为应届毕业生，故不再考虑此因素。

综上所述，本文所确定的求职者意愿模型为：

Job\_seeker\_inten (ID, T, M, GA, GF, G, GP)

其中，ID：求职者编号；

T： 到岗时间；

M： 期望薪水；

GA： 工作地点；

GF： 工作性质；

G： 工作行业；

GP： 工作职位

### 3.3 求职者职位竞争力评价模型

#### 3.3.1 定义

求职者职位竞争力评价模型：用于评价投递同一职位的求职者之间的竞争力。

#### 3.3.2 作用

通过建立求职者职位竞争力评价模型，为网络招聘系统引入辅助决策机制，

为网络求职者投递简历提供参考，避免投递简历的盲目性和随意性。

### 3.3.3 模型构建方法

首先，确定评价指标。所谓的评价指标是指从网络招聘者角度评价，影响网络求职者获得网络招聘者面试机会的因素。因此，我们要以企业的职位需求为核心，综合考虑求职者的其它因素的原则来确定指标。根据这个原则，我们建立 N+1 评价指标模型。其中“1”指职位需求，“N”指“个人基本情况”、“教育背景”、“工作经验”、“获奖情况”、“自我评价”、“社会实践”、“专业技能”、“求职意愿”和“语言能力”等方面的信息。

其次，确定指标评判方法。首先要分析问题域特征，了解目前常用的解决此类问题的方法，以及它们的优缺点，再具体介绍了本文的基于优势关系的粗集多属性排序方法及其在网络招聘信息个性化推荐中的应用。

#### (1) 评价指标确定

根据评价指标确定原则，考虑到数据的来源及网络招聘中数据项量化问题，本文对“工作经验”、“自我评价”、“社会实践”和“专业技能”等带有较强主观色彩的指标不作比较，网络招聘者可在后绪的面试和笔试中进行进一步的考查和筛选。

因此，本文建立以职位信息为核心，综合考虑个人基础信息、教育背景、获奖情况、语言能力的 4+1 评价指标模型，如下所示：

$Job\_seeker\_Com\_Eva (ID, GN (EB\_E, M, GA, GF, LQ\_L, S), PI (PI\_P), EB (EB\_E, EB\_S), AW (AW\_L), LQ (LQ\_L))$

其中，ID：求职者编号；

GN：职位需求；

EB\_E：学历；

LQ\_L：语言；

M：专业；

GA：工作地点；

GF：工作职位；

S：性别；

PI: 个人基本情况;

PI\_P : 政治面貌;

EB: 教育背景;

EB\_E: 学历;

EB\_S: 学校;

AW: 获奖情况;

AW\_L 奖项级别;

LQ: 语言能力;

LQ\_L : 语言;

## (2) 评判方法确定

### 1) 问题域特征分析

该评价模型具有以下几个特征, 分别是:

①. 大多数情况下每个职位都有多个求职者, 因此这是一个在多方案中择优问题;

②. 对求职者竞争力的影响因素比较多, 如学历、毕业院校、工作经验等, 因此具有多属性决策问题的特征;

③. 人与人之间比较, 是一个复杂、模糊、难定量的问题;

④. 在多个求职者之间进行竞争力排序, 必须两两之间都进行比较, 才能保证排序的合理性和正确性。

基于以上分析可知, 评价网络求职者的职位竞争力是个多属性决策问题<sup>[43]</sup> (MADM 问题), 多属性决策问题的特征是, 要有多个属性准则的制约, 并且各个属性之间常常会遇到冲突的状况, 以及各属性度量的量纲一般是不相同的<sup>[44,45]</sup>。

### 2) 问题域的国内外研究现状

目前多属性决策问题解决方法主要有以下几种, 分别是:

①. 层次分析法<sup>[46,47]</sup>

②. 多属性效用理论<sup>[48]</sup>

③. LINMAP 方法<sup>[49,50]</sup>

④. 模糊综合评判法<sup>[51]</sup>



这些方法计算方式各不相同,应用领域也不大相同,但大多都通过计算各个方案的属性的权重来实现,因此本质还是停留在权重的确定上。而本文采用基于优势关系的粗集多属性排序方法<sup>[52]</sup>,是建立在偏序关系的基础上,可有效避免权重的计算。下一小节重点介绍基于优势关系的粗集多属性排序方法。

### (3) 基于优势关系的粗集多属性竞争力排序方法

1) 该方法的主要思想是:以分类为基础,比如有一方案 A,将比方案 A 好的归为一类,比 A 差的归为一类;根据粗集,利用包含度<sup>[53,54]</sup>(包含度主要是描述一个模糊集包含在另一个模糊集中的程度)计算,在某一个指标条件下,两个方案的优劣程度;最后再用算术平均值作为排序的综合指数,进行排序<sup>[52]</sup>。

具体算法如下:

第一步:建立有序信息系统  $(X, Y, F)$ , 其中  $X = \{x_1, x_2, \dots, x_n\}$  为非空的有限集,表示方案集;  $Y = \{y_1, y_2, \dots, y_n\}$  为非空的属性集;  $F = \{f_k : X \rightarrow V_k (k \leq n)\}$  为方案与属性之间的关系集,  $V_k$  是属性  $a_i$  的有限值域且是偏序集,  $B \subset Y$ ;

第二步:根据次占优关系计算各方案的优劣类;

$$[x_i]_B^{\leq} = \{x_j | (x_i, x_j) \in R_B^{\leq}\} = \{x_j | f_k(x_i) \leq f_k(x_j) (a_k \in B)\} \quad (3.1)$$

第三步:根据包含度计算两两方案之间的优势度;

$$R_B(x_i, x_j) = \frac{|([x_i]_B^{\leq}) \cap [x_j]_B^{\leq}|}{|X|} \quad (3.2)$$

第四步:计算综合优势度;

$$R_B(x_i) = \frac{1}{n-1} \sum_{j \neq i} R_B(x_i, x_j) \quad (3.3)$$

第五步:根据综合优势度进行排序。

## 2) 实例

### (1) 求职者职位竞争力评价信息系统

由本小节可知,求职者职位竞争力评价指标集为  $Y = \{\text{政治面貌, 毕业院校, 学历, 学生奖励, 外语水平}\} = \{a_1, a_2, \dots, a_5\}$ , 求职者集合  $X = \{x_1, x_2, \dots, x_{10}\}$ , 其中  $x_1$  为当前求职者,  $x_2, \dots, x_{10}$  为已经投递简历的求职者。对每一个求职者  $x_i$  和每个评价指标  $a_j$  都有指标值  $f_{a_j}(x_i)$  (如表 3.1 所示)与之对应,  $i=1, 2, \dots, 5$ ;  $j=1, 2, \dots, 10$ 。

由此我们得到一个求职者竞争力评价模型的信息系统(X,Y,F)。

表 3.1 10 个求职者的评价指标信息  
Tab.3.1 Evaluation information of 10 job-seekers

求职者	政治面貌	毕业院校	学历	学生奖项	外语水平
X <sub>1</sub>	党员	28	硕士	院级	六级
X <sub>2</sub>	团员	5	本科	国家级	四级
X <sub>3</sub>	党员	36	大专	省级	六级
X <sub>4</sub>	团员	10	硕士	院级	六级
X <sub>5</sub>	党员	4	本科	国家级	六级
X <sub>6</sub>	团员	55	大专	省级	托业
X <sub>7</sub>	团员	30	硕士	校级	托业
X <sub>8</sub>	团员	8	本科	院级	托福
X <sub>9</sub>	团员	60	大专	省级	六级
X <sub>10</sub>	党员	25	硕士	校级	六级

(2)改进的基于优势关系的多属性排序法

①. 将政治面貌、学历、学生奖项和外语水平按其本身所固有的客观属性进行量化，如毕业院校按 2007 年专业排名量化，以简化研究，具体量化值如表 3.2 所示：

表 3.2 评价指标量化值  
Tab.3.2 The value of evaluation index

分类	具体因素	量化值				
		5	4	3	2	1
个人基本情况	政治面貌	党员	团员	群众		
教育背景	毕业院校	1~20	21~40	41~60	61~80	81~100
	学历	博士	硕士	本科	大专	
在校情况	学生奖项	国家级	省级	校级	院级	无
语言能力	外语水平	托福雅思	托业	六级	四级	

②. 对于次占优关系，本文按照多数原则，以 5 个指标中有 4 个指标占优即为占优的准则来进行排序。

(3)求职者竞争力计算及排序过程

a)建立 10 个求职者的评价指标量化信息，如下表 3.3 所示：

表 3.3 10 个求职者的竞争指标量化信息  
Tab.3.3 The competitive index of 10 job seekers

求职者	政治面貌	毕业院校	学历	学生奖项	外语水平
X <sub>1</sub>	5	4	4	2	3
X <sub>2</sub>	4	3	3	5	2
X <sub>3</sub>	5	4	2	4	3
X <sub>4</sub>	4	5	4	3	3
X <sub>5</sub>	5	5	3	5	3
X <sub>6</sub>	4	3	2	4	4
X <sub>7</sub>	4	4	4	4	4
X <sub>8</sub>	4	5	3	2	5
X <sub>9</sub>	4	3	2	4	3
X <sub>10</sub>	5	4	4	3	3

b) 根据次占优关系，计算各求职者优势类，其结果如下：

$[X_1]_{Rc} = \{ X_1, X_3, X_4, X_5, X_7, X_{10} \};$   
 $[X_2]_{Rc} = \{ X_1, X_2, X_4, X_5, X_7, X_8, X_{10} \};$   
 $[X_3]_{Rc} = \{ X_1, X_3, X_5, X_7, X_{10} \};$   
 $[X_4]_{Rc} = \{ X_4, X_5, X_7, X_{10} \};$   
 $[X_5]_{Rc} = \{ X_5, \};$   
 $[X_6]_{Rc} = \{ X_2, X_3, X_5, X_6, X_7, X_8, X_9 \};$   
 $[X_7]_{Rc} = \{ X_7 \};$   
 $[X_8]_{Rc} = \{ X_4, X_5, X_8 \};$   
 $[X_9]_{Rc} = \{ X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10} \};$   
 $[X_{10}]_{Rc} = \{ X_1, X_3, X_4, X_5, X_7, X_{10} \};$

相应的补集为：

$$[x_1]_{R_c}^c = \{ x_2, x_6, x_8, x_9 \};$$

$$[x_2]_{R_c}^c = \{ x_3, x_6, x_9 \};$$

$$[x_3]_{R_c}^c = \{ x_2, x_4, x_6, x_8, x_9, \};$$

$$[x_4]_{R_c}^c = \{ x_1, x_2, x_3, x_6, x_8, x_9 \};$$

$$[x_5]_{R_c}^c = \{ x_1, x_2, x_3, x_4, x_6, x_7, x_8, x_9, x_{10} \};$$

$$[x_6]_{R_c}^c = \{ x_1, x_4, x_{10} \};$$

$$[x_7]_{R_c}^c = \{ x_1, x_2, x_3, x_4, x_5, x_6, x_8, x_9, x_{10} \};$$

$$[x_8]_{R_c}^c = \{ x_1, x_2, x_3, x_6, x_7, x_9, x_{10} \};$$

$$[x_9]_{R_c}^c = \emptyset;$$

$$[x_{10}]_{R_c}^c = \{ x_2, x_6, x_8, x_9 \};$$

c)根据公式 (3.4) 计算优势度

$$R(x_i, x_j) = \frac{|([x_i]_{R_c}^c) \cup ([x_j]_{R_c}^c)|}{|x|} \quad (3.4)$$

记优势度向量为  $\vec{X}_{x_i}$ , 简记为  $\vec{X}_i$

$$\vec{X}_i = \{ R(x_i, x_1), R(x_i, x_2), \dots, R(x_i, x_{i-1}), R(x_i, x_i), \dots, R(x_i, x_{10}) \}$$

$i=1,2,\dots,10$ ,根据公式(3.4)可得计算结果为:

$$\vec{X}_1 = \frac{1}{10} \{ 9, 9, 8, 5, 7, 5, 6, 10, 10 \};$$

$$\vec{X}_2 = \frac{1}{10} \{ 8, 7, 7, 4, 7, 4, 6, 10, 8 \};$$

$$\vec{x}_3 = \frac{1}{10} \{ 10, 9, 8, 6, 8, 6, 6, 10, 10 \};$$

$$\vec{x}_4 = \frac{1}{10} \{ 10, 10, 9, 7, 8, 7, 8, 10, 10 \};$$

$$\vec{x}_5 = \frac{1}{10} \{ 10, 10, 10, 10, 10, 9, 10, 10, 10 \};$$

$$\vec{x}_6 = \frac{1}{10} \{ 6, 7, 6, 5, 4, 4, 5, 10, 6 \};$$

$$\vec{x}_7 = \frac{1}{10} \{ 10, 10, 10, 10, 9, 10, 9, 10, 10 \};$$

$$\vec{x}_8 = \frac{1}{10} \{ 9, 10, 8, 9, 8, 9, 7, 10, 9 \};$$

$$\vec{x}_9 = \frac{1}{10} \{ 6, 7, 5, 4, 1, 7, 1, 3, 6 \};$$

$$\vec{x}_{10} = \frac{1}{10} \{ 10, 9, 9, 8, 5, 7, 5, 6, 10 \};$$

d) 根据公式 (3.5) 计算综合优势度

$$R(x_i) = \frac{1}{|X|-1} \sum_{i \neq j} R(x_i, x_j) \quad (3.5)$$

所得结果为:

$$R(x_1) = \frac{69}{10 \times 9}$$

$$R(x_2) = \frac{61}{10 \times 9}$$

$$R(x_3) = \frac{73}{10 \times 9}$$

$$R(x_4) = \frac{79}{10 \times 9}$$

$$R(x_5) = \frac{89}{10 \times 9}$$

$$R(x_6) = \frac{53}{10 \times 9}$$

$$R(x_7) = \frac{88}{10 \times 9}$$

$$R(x_8) = \frac{79}{10 \times 9}$$

$$R(x_9) = \frac{40}{10 \times 9}$$

$$R(x_{10}) = \frac{69}{10 \times 9}$$

e) 根据综合优势度可得到 10 个求职者的竞争力排序结果:

$$x_5 > x_7 > x_4 \approx x_8 > x_3 > x_1 \approx x_{10} > x_2 > x_6 > x_9$$

f) 结果分析

从求职者竞争力排序结果可看出, 第 5 个求职者的竞争力最强, 当前求职者

位于中间位置。求职者可以根据参考信息，再根据自己的喜好决定是否投递简历。

### 3.4 求职者职位适合度评价模型

#### 3.4.1 定义

求职者职位适合度评价模型，是对同一求职者，评价多个职位信息的适合度。

#### 3.4.2 作用

解决网络求职招聘中信息巨量性、海量性引发的信息迷航问题，为求职者准确投递简历提供辅助决策机制。

#### 3.4.3 模型构建方法

求职者职位适合度评价模型是对某个求职者，评价多个职位信息的适合程度，而求职者职位竞争力评价模型是对某个招聘者，评价多个求职者信息，因此二者本质是一样的，因此本模型也采用与求职者职位竞争力评价模型相同的解决思路与算法，在此不再赘述。

##### (1) 评价指标的确定

经分析最终确定以网络求职者的求职意愿为核心，综合考虑网络招聘者的其它因素，建立了一个 2+1 的适合度评价指标体系。

$Job\_seeker\_fit\_Eva(Job\_seeker\_inten, Employ\_inf(Com\_name, Com\_size), Job\_seeker\_inf(Emp\_count, EB\_E, LQ\_L, M))$

其中， $Job\_seeker\_inten$ ：求职意愿

$Employ\_inf$ ：网络招聘者基本信息

$Com\_name$ ：公司名称

$Com\_size$ ：公司规模

$Job\_seeker\_inf$ ：对求职者要求信息

$Emp\_count$ ：招聘人数

$EB\_E$ ：学历要求

$LQ\_L$ ：语言能力

$M$ ：职位月薪

##### (2) 职位适合度评价实例

1)求职者职位适合度评价信息系统

由本小节可知，职位适合度评价指标集为  $Y=\{\text{公司名称, 公司规模, 招聘人数, 学历要求, 语言能力, 职位月薪}\}=\{a_1,a_2, \dots a_6\}$ ，职位集合  $X=\{x_1, x_2, \dots, x_{10}\}$ ，对每一个职位  $x_i$  和每个评价指标  $a_j$  都有指标值  $f_{a_j}(x_i)$  (如表 3.4 所示)与之对应， $i=1,2,\dots,6$ ;  $j=1,2,\dots,10$ . 由此我们得到一个职位适合度评价模型的信息系统  $(X,Y,F)$ 。

表 3.4 10 个职位的适合度评价指标信息

Tab.3.4 10 The fitness index values of 10 posts

职位	公司名称	公司规模	招聘人数	学历要求	语言能力	职位月薪
$x_1$	百度	10255	100	本科以上	六级	9000
$x_2$	google	9520	85	硕士以上	六级	7500
$x_3$	中兴	1420	73	本科以上	六级	7000
$x_4$	大唐	850	20	大专以上	四级	3500
$x_5$	金山在线	320	35	硕士以上	四级	6000
$x_6$	东软	800	46	本科以上	四级	4500
$x_7$	中国联通	10480	100	硕士以上	六级	4600
$x_8$	奥美电子	300	30	本科以上	四级	4000
$x_9$	中兴通讯	450	55	本科以上	四级	4200
$x_{10}$	四川托日	90	20	大专以上	四级	2500

2)改进的基于优势关系的多属性排序法

①. 将各指标按其本身所固有的客观属性进行量化，以简化研究。具体量化值如下表 3.5 所示：

表 3.5 职位适合度评价指标量化标准表  
Tab.3.5 The index value of post fitness evaluation

评价指标	量化值									
	10	9	8	7	6	5	4	3	2	1
公司名称	1~10	11~20	21~30	31~40	41~50	51~60	61~70	71~80	81~90	>90
公司规模	>9999	9000~	7000~	5000~	3000~	1000~	500~	100~	50~	<50
		9999	8999	6999	4999	2999	999	499	99	
招聘人数	>100	90~	80~8	70~	60~	50~	40~	30~	20~	<20
		99	9	79	69	59	49	39	29	
学历要求	高中		大专		本科		硕士		博士	
语言能力	无		四级		六级		托业		托福 雅思	
职位月薪	>9999	9000~	8000~	7000~	6000~	5000~	4000~	3000~	2000~	<200
		9999	8999	7999	6999	5999	4999	3999	2999	0

其中，“公司名称”根据其 2007 年企业排名来量化，“公司规模”“招聘人数”根据各大招聘网站的分类方法来量化，“学历要求”“语言能力”根据一般的社会常识分类进行量化，“职位月薪”根据 2008 年网络统计的大学生和研究生毕业的平均工资来量化。

②. 对于次占优关系，本文按照多数原则，以 6 个指标中有 5 个指标占优即为占优的准则来进行排序。

(3)职位适合度计算及排序过程

①. 根据上一小节中表 3.5 职位适合度评价指标量化表来建立 10 个职位的适合度评价指标量化信息，如表 3.6 所示：



表 3.6 10 个职位的适合度评价指标量化信息

Tab. 3.6 The fitness index of 10 posts

职位	公司名称	公司规模	招聘人数	学历要求	语言能力	职位月薪
X <sub>1</sub>	10	10	10	6	6	9
X <sub>2</sub>	10	9	8	4	6	7
X <sub>3</sub>	5	5	7	4	6	7
X <sub>4</sub>	7	4	2	8	8	3
X <sub>5</sub>	6	3	3	4	8	6
X <sub>6</sub>	4	4	4	6	6	4
X <sub>7</sub>	8	10	10	4	4	4
X <sub>8</sub>	2	3	3	6	6	4
X <sub>9</sub>	7	3	5	6	6	4
X <sub>10</sub>	1	2	2	8	6	2

②. 基于次占优关系，根据优势类公式计算各求职者优势类，其结果如下：

- [X<sub>1</sub>] R<sub>c</sub>={ X<sub>1</sub>};
- [X<sub>2</sub>] R<sub>c</sub>={ X<sub>1</sub>, X<sub>2</sub>};
- [X<sub>3</sub>] R<sub>c</sub>={ X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>};
- [X<sub>4</sub>] R<sub>c</sub>={ X<sub>4</sub> };
- [X<sub>5</sub>] R<sub>c</sub>={ X<sub>1</sub>, X<sub>2</sub>, X<sub>5</sub>};
- [X<sub>6</sub>] R<sub>c</sub>={ X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>, X<sub>6</sub>, X<sub>9</sub>};
- [X<sub>7</sub>] R<sub>c</sub>={ X<sub>1</sub>, X<sub>7</sub>};
- [X<sub>8</sub>] R<sub>c</sub>={ X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>, X<sub>5</sub>, X<sub>6</sub>, X<sub>8</sub>, X<sub>9</sub> };
- [X<sub>9</sub>] R<sub>c</sub>={ X<sub>1</sub>, X<sub>2</sub>, X<sub>9</sub>};
- [X<sub>10</sub>] R<sub>c</sub>={ X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>, X<sub>4</sub>, X<sub>5</sub>, X<sub>6</sub>, X<sub>8</sub>, X<sub>9</sub>, X<sub>10</sub> };

相应的补集为：

- [X<sub>1</sub>]<sup>c</sup><sub>R<sub>c</sub></sub>={ X<sub>2</sub>, X<sub>3</sub>, X<sub>4</sub>, X<sub>5</sub>, X<sub>6</sub>, X<sub>7</sub>, X<sub>8</sub>, X<sub>9</sub>, X<sub>10</sub> };
- [X<sub>2</sub>]<sup>c</sup><sub>R<sub>c</sub></sub>={ X<sub>3</sub>, X<sub>4</sub>, X<sub>5</sub>, X<sub>6</sub>, X<sub>7</sub>, X<sub>8</sub>, X<sub>9</sub>, X<sub>10</sub> };
- [X<sub>3</sub>]<sup>c</sup><sub>R<sub>c</sub></sub>={ X<sub>4</sub>, X<sub>5</sub>, X<sub>6</sub>, X<sub>7</sub>, X<sub>8</sub>, X<sub>9</sub>, X<sub>10</sub> };

$$[x_4]_{R_c}^c = \{ x_1, x_2, x_3, x_5, x_6, x_7, x_8, x_9, x_{10} \};$$

$$[x_5]_{R_c}^c = \{ x_3, x_4, x_6, x_7, x_8, x_9, x_{10} \};$$

$$[x_6]_{R_c}^c = \{ x_4, x_5, x_7, x_8, x_{10} \};$$

$$[x_7]_{R_c}^c = \{ x_2, x_3, x_4, x_5, x_6, x_8, x_9, x_{10} \};$$

$$[x_8]_{R_c}^c = \{ x_4, x_7, x_{10} \};$$

$$[x_9]_{R_c}^c = \{ x_3, x_4, x_5, x_6, x_7, x_8, x_{10} \}$$

$$[x_{10}]_{R_c}^c = \{ x_7 \};$$

③. 根据公式 (3.6) 计算优势度

$$R(x_i, x_j) = \frac{|([x_i]_{R_c}^c) \cup ([x_j]_{R_c}^c)|}{|x_i|} \quad (3.6)$$

记优势度向量为  $\vec{X}_{x_i}$ , 简记为  $\vec{X}_i$

$$\vec{X}_i = \{ R(x_i, x_1), R(x_i, x_2), \dots, R(x_i, x_{i-1}), R(x_i, x_i), \dots, R(x_i, x_{10}) \}$$

$i=1,2,\dots,10$ , 根据公式(3.6)可得计算结果为:

$$\vec{X}_1 = \frac{1}{10} \{ 10, 10, 9, 10, 10, 10, 10, 10, 10, 10 \};$$

$$\vec{X}_2 = \frac{1}{10} \{ 9, 10, 8, 10, 10, 9, 10, 10, 10, 10 \};$$

$$\vec{X}_3 = \frac{1}{10} \{ 8, 9, 7, 9, 10, 8, 10, 9, 10, 10 \};$$

$$\vec{X}_4 = \frac{1}{10} \{ 9, 9, 9, 9, 9, 9, 9, 9, 10, 10 \};$$

$$\vec{X}_5 = \frac{1}{10} \{ 8, 9, 9, 7, 9, 8, 10, 9, 10, 10 \};$$

$$\vec{X}_6 = \frac{1}{10} \{ 6, 7, 8, 5, 7, 6, 10, 8, 10, 10 \};$$

$$\vec{x}_7 = \frac{1}{10} \{ 9, 9, 9, 8, 9, 9, 9, 9, 9 \};$$

$$\vec{x}_8 = \frac{1}{10} \{ 4, 5, 6, 3, 6, 8, 4, 6, 10 \};$$

$$\vec{x}_9 = \frac{1}{10} \{ 8, 9, 9, 7, 9, 10, 8, 10, 10 \};$$

$$\vec{x}_{10} = \frac{1}{10} \{ 2, 3, 4, 2, 4, 6, 2, 8, 4 \};$$

④. 根据公式 (3.7) 计算综合优势度

$$R(x_i) = \frac{1}{|x_i|-1} \sum_{i \neq j} R(x_i, x_j) \quad (3.7)$$

所得结果为:

$$R(x_1) = \frac{89}{10 \times 9}$$

$$R(x_2) = \frac{86}{10 \times 9}$$

$$R(x_3) = \frac{80}{10 \times 9}$$

$$R(x_4) = \frac{82}{10 \times 9}$$

$$R(x_5) = \frac{79}{10 \times 9}$$

$$R(x_6) = \frac{67}{10 \times 9}$$

$$R(x_7) = \frac{80}{10 \times 9}$$

$$R(x_8) = \frac{52}{10 \times 9}$$

$$R(x_9) = \frac{80}{10 \times 9}$$

$$R(x_{10}) = \frac{35}{10 \times 9}$$

⑤. 根据综合优势度可得到 10 个职位的适合度排序结果:

$$x_1 > x_2 > x_4 > x_3 \approx x_7 \approx x_9 > x_5 > x_6 > x_8 > x_{10}$$

⑥. 结果分析

从排序结果可以看出第一个职位的适合度比较高,而第 10 个职位的适合度比较低.

### 3.5 小结

本章从求职的定义 3 和定义 4 出发, 分别介绍了网络求职招聘过程模型、求职者求职意愿模型、求职者职位竞争力评价模型和求职者职位适合度评价模型的定义、作用、建立原则及方法。

## 第 4 章 原型系统设计与实现

本章主要从系统总体设计、数据库设计和求职者功能设计几个方面来介绍网络求职个性化信息推荐的具体实现。

### 4.1 系统总体设计

本文设计并实现了网络招聘信息个性化推荐系统，为了下文表述方便，将其简称为 IJRS (Internet Job Recommendation System)。

IJRS 总体设计非常简单，需求和功能都相当明确，本节详细介绍 IJRS 网站的系统功能、结构及数据库设计等方面。

#### 4.1.1 系统需求

IJRS 与目前的网络求职网站不同，以“求职者过程模型”为依据，实现基于求职者求职意愿模型、求职者职位竞争力评价模型、求职者职位适合度评价模型的网络招聘信息个性化推荐服务。

在实际设计中，IJRS 基于 C/S 架构，在 Visual studio 2005 平台下，以 C#为开发语言，SQL server 2005 作为数据库。

#### 4.1.2 系统功能

在 IJRS 中共有三种类型的角色及一个智能体 Agent，三种类型的角色分别是系统管理员、求职者和招聘者。系统管理员主要进行一系列的管理操作，而主要功能由求职者和招聘者来完成，图 4.1 的(a), (b), (c)分别为 IJRS 中两大角色（求职者和招聘者）及智能体 Agent 的功能结构图。

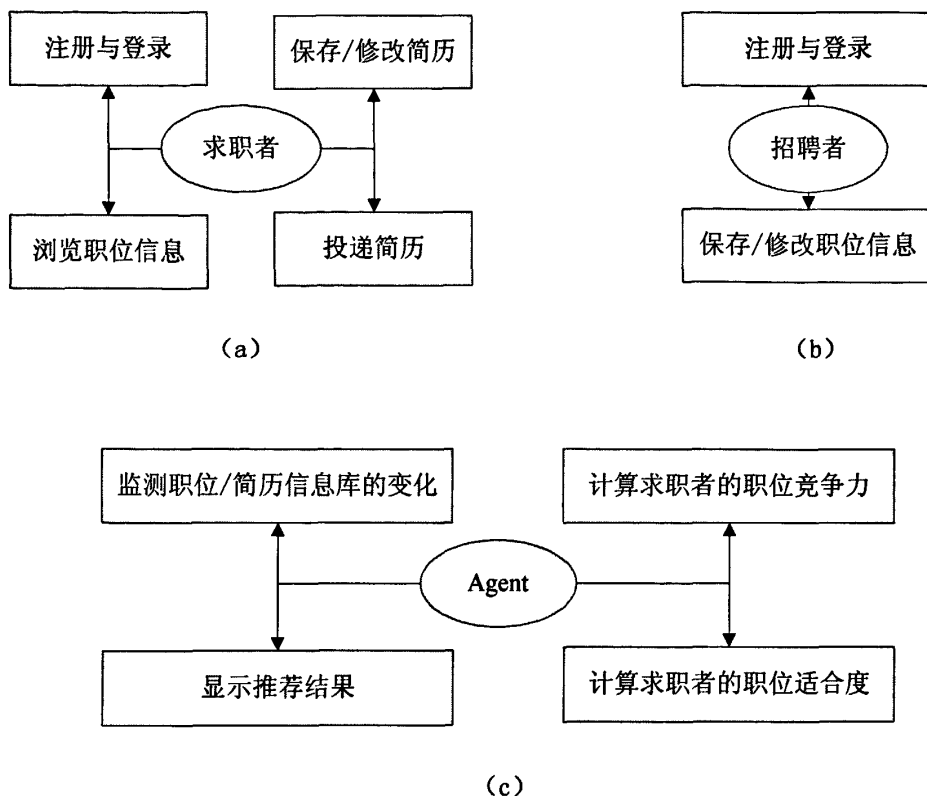


图 4.1 功能结构图

Fig 4.1 function structure chart

作为一个求职者，主要完成以下操作：

- (1) 注册到 IJRS 站点
- (2) 保存简历
- (3) 修改简历
- (4) 浏览招聘信息
- (5) 投递简历

作为一个招聘者，可以进行如下的操作：

- (1) 注册到 IJRS 站点
- (2) 发布一个或多个职位信息
- (3) 修改职位信息

智能体 Agent 完成的功能主要如下：

- (1) 监测职位信息库的变化
- (2) 监测简历信息库的变化
- (3) 计算求职者竞争力
- (4) 计算求职者职位适合度
- (5) 显示推荐结果

系统总体流程图如图 4.2 所示。

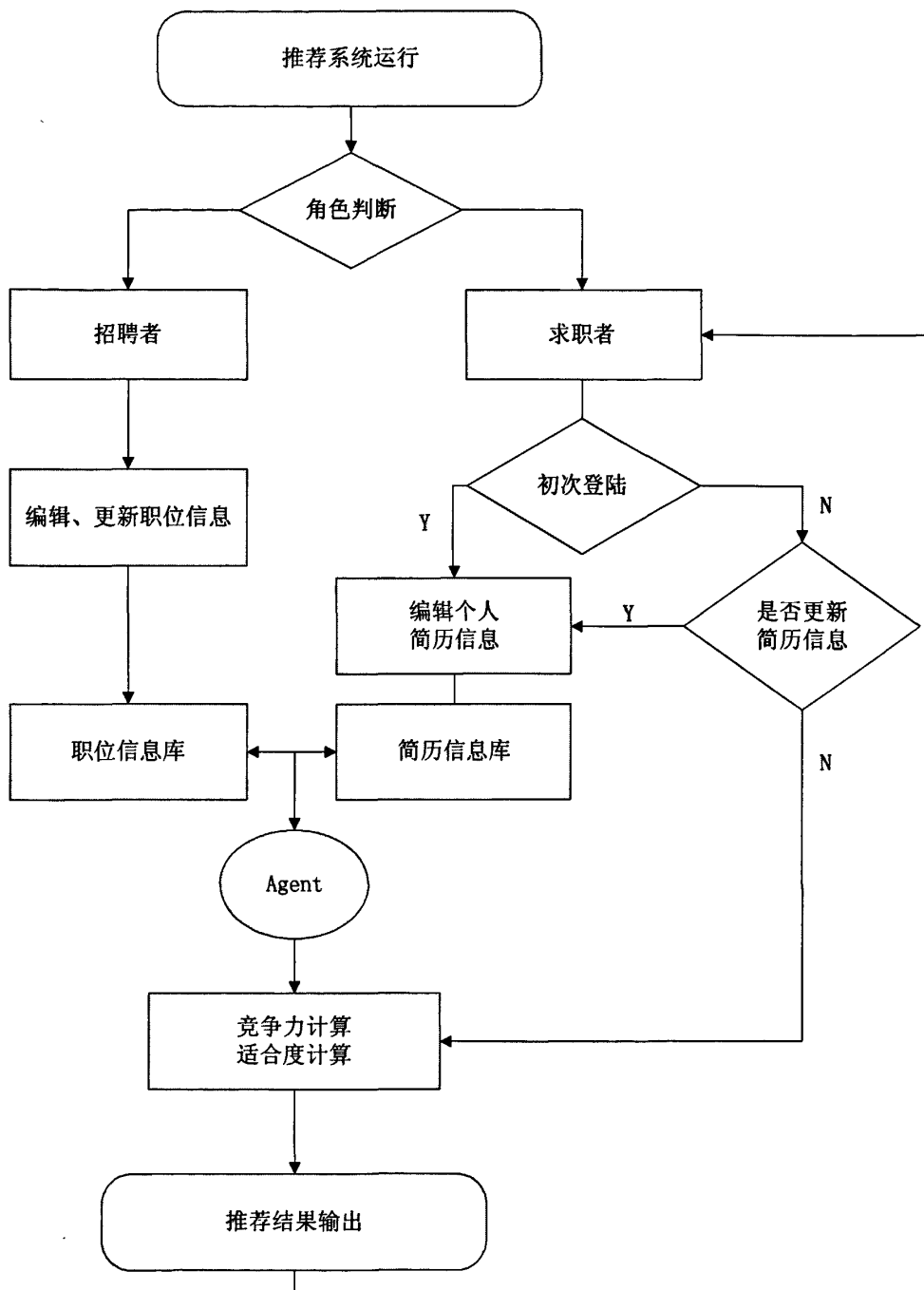


图 4.2 IJRS 系统流程图

Fig 4.2 IJRS system flow chart

系统开始运行时，首先判断是求职者还是招聘者，如果是求职者，再判断是

否为初次登陆的求职者，如果是初次登陆，则必须先注册个人基本信息，然后保存到简历库中，智能体 Agent 会根据从简历库中抽取“求职意愿模型”“求职者职位竞争力评价模型”所需要的信息，从职位信息库中抽取“求职者职位适合度评价模型”所需要的信息，并进行计算，最后按照求职者要求进行排序，并将结果推荐给求职者。

如果是招聘者，可以对职位进行增加、删除和修改等操作。

智能体 Agent 会每隔一个时间单位对简历信息库和职位信息库进行维护，查看是否有更新，如果有更新，则重新计算竞争力和适合度，并将满足求职者要求的职位信息推荐给求职者。

4.2 数据库设计

IJRS 网站使用 SQL Server 2005 作为后端数据库，在 Visual Studio 2005（以下简称 VS2005）中直接建立 IJRSDB.mdf 文件，IJRS 中共有 7 个表，本文将这些表分为静态表和动态表。

静态表是指在系统初始化时由管理员进行输入的，一般较少变动，所以本文称之为静态表。在 IJRS 中属于静态表的如表 4.1 所示。

表 4.1 静态表  
Tab.4.1 Static table

表名称	说明
IJRSDB_ComChange	表示竞争力指标的量化表
IJRSDB_FitChange	表示适合度指标的量化表

动态表是指在与系统交互过程中产生的，比如求职者录入的简历信息，则在简历表中产生了一条记录。IJRS 中属于动态表的如表 4.2 所示。



表 4.2 动态表  
Tab.4.2 Dynamic table

表名称	说明
IJRSDB_Resumes	表示求职者的简历表
IJRSDB_JobPosittings	表示招聘者的招聘职位信息表
IJRSDB_Companies	表示招聘者的公司信息
IJRSDB_Com Property	表示抽取的竞争力指标集
IJRSDB_FitProperty	表示抽取的适合度指标集

4.3 功能模块设计

软件设计包括系统输入、分析与输出。因此，本小节主要从这三个方面来介绍，职位信息个性化推荐系统结构图如图 4.3 所示：

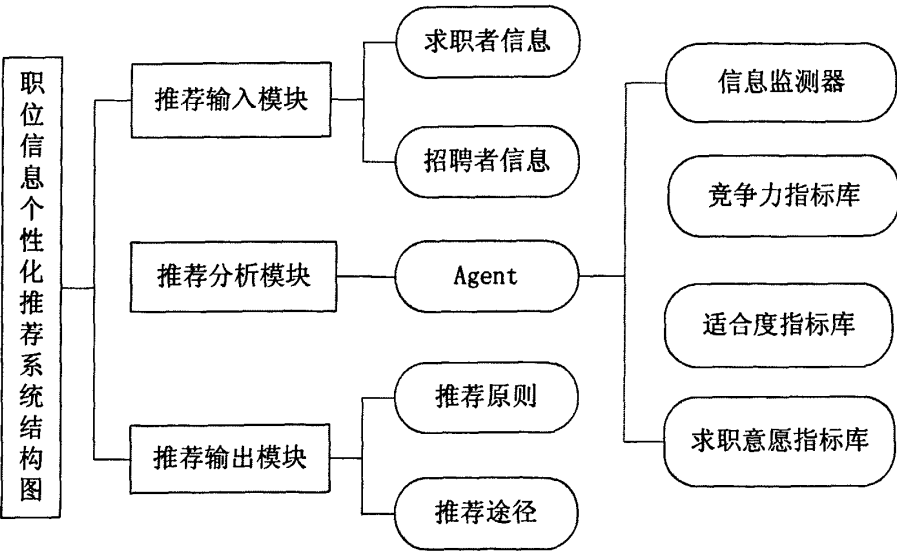


Fig 4.3 Job information individuation recommend system structure chart

4.3.1 系统输入

(1) 求职者模块设计

求职者的访问网站时主要进行以下 5 种操作：

1) 注册到 IJRS 站点：只有初次登录时，进行注册。

2) 保存简历：

3) 修改简历

4) 浏览招聘信息

5) 投递简历

## (2) 招聘者模块设计

招聘者的访问网站时主要进行以下 4 种操作：

1) 注册到 IJRS 站点

2) 发布一个或多个职位信息

3) 修改职位信息

同样，注册职位信息，也只设计了后续需要的指标。

### 4.3.2 系统分析

本模块主要包括信息监测器、求职者求职意愿库、求职者竞争力指标库及求职者职位适合度指标库。

信息监测器的功能是动态监测简历库和职位库，以便及时更新竞争力指标库和职位适合度指标库，并计算竞争力和适合度，将新信息传递给求职者。

#### (1) 信息监测器

功能：信息监测器采用“连续查询”的策略，对求职者个人信息、职位信息进行实时监测，并采用相应的解决方法。

如果监测到有职位信息更新，则先根据当前求职者的求职意愿模型判断其职位是否满足要求，如果满足要求，则抽取职位适合度模型所需要的指标，加入到职位适合度指标库中，同时移出已经投递简历的职位信息，对新的职位指标库计算其职位适合度和求职者的竞争力。如果没有职位更新，经过一定时间段后，将当前求职者已经投递简历的职位信息移出即可。

如果监测到求职者更新了求职意愿指标中的属性，则按照求职者的求职意愿对职位重新进行过滤，对符合要求的职位信息抽取“求职者职位适合度评价模型”所需要的指标信息，重新建立适合度指标库，计算求职者的职位适合度和求职者的竞争力。

## (2) 计算竞争力

功能：根据基于优势关系的多属性决策方法计算求职者的职位竞争力。

## (3) 计算适合度

功能：根据基于优势关系的多属性决策方法计算求职者的职位适合度。

### 4.3.3 系统输出

功能：根据求职者的喜好，选择一种推荐原则对职位信息进行重新排序，并将推荐结果显示给求职者。

本论文的推荐原则主要有以下两个：

(1) 根据求职者的职位竞争力进行排序；

(2) 根据求职者的职位适合度进行排序。

推荐途径包括邮件、手机短信等各种通讯设备。

## 4.4 网络求职招聘与个性化推荐实例

### (1) 系统已知条件

假设条件如下：

1) 职位库中有 10 个职位信息及每个职位已经投递简历的求职者数量及编号，职位信息与表 3.4 相同。

2) 系统的简历库中有 10 个求职者的求职信息，求职者信息与表 3.1 相同。

3) ID 为 1 的求职者（记为 X1）为当前求职者，已经注册完简历信息，如图 4.4 所示：

图 4.4 填写简历示例图

Fig.4.4 The sample of resume registration

4) 由于目前的人才招聘网站的信息过滤技术已经相当成熟, 因此在实际操作中本文仍然使用已有的信息过滤技术, 本文的 IJRS 在双方都互相满足的基础上进行职位信息推荐。

目标:

根据本文建立的求职者求职意愿模型、求职者职位竞争力评价模型、求职者职位适合度评价模型为其推荐个性化的职位信息。

## (2) 解决方法

具体步骤如下:

### 1) Agent 监测到简历库中的信息变化

当 Agent 监测到简历库中的信息变化时, 从简历库中重新抽取求职者竞争力模型需要的指标, 存入表 IJRSDB\_ComProperty 中, 以便计算竞争力, 此时职位库并没有发生变化, 因此不需要重新抽取职位适合度指标集。

2) 根据模型算法先在整个职位集上计算职位适合度指数, 再对每个职位计算当前求职者竞争力指数, 此时, 求职者可以根据竞争力排名和职位适合度排名, 查看职位的具体信息并投递简历。如图 4.5 所示:

网络招聘信息个性化推荐系统  
Intelligent Technology Research Center

欢迎你! 张一!  
[注销]

推荐结果

选择	公司名称	职位名称	工作地点	竞争力排名	适合度排名	已投递简历数
选择	中国联通	软件测试	山西	7	1	8
选择	金山在线	软件工程师	大连	5	2	6
选择	大唐	软件测试	大连	4	3	7
选择	中兴	软件工程师	大连	3	4	5
选择	百度	软件工程师	北京	1	5	4
选择	东软	软件工程师	大连	6	6	9
选择	google	软件测试	大连	2	7	3
选择	四川托日	软件工程师	四川	10	8	10
选择	中兴通讯	软件测试	深圳	9	9	10
选择	奥美电子	软件测试	深圳	8	10	9

图 4.5 推荐结果示例图

Fig.4.5 The sample of recommendation result

### 3)发布新的职位信息

招聘者新发布了一个职位信息，编号为 11，并且此时求职者处于离线状态，当 Agent 监测到职位信息变化时，从职位库中重新抽取职位适合度指标集，但是对求职者已经投递简历的职位信息如编号为 1 的职位，则不需要再抽取信息，在新的指标库上重新计算职位适合度和竞争力后将结果推荐给求职者。

由于此时求职者处于离线状态，为了不错过可能的机会，用邮箱或手机短信的方式通知求职者有新的适合的职位信息。

## 4.5 小结

本章首先介绍了 IJRS 的系统总体设计与数据库设计，接着介绍了 IJRS 的各个模块功能，最后通过实例详细介绍了如何实现职位信息的个性化推荐。

## 第 5 章 结论与展望

### 5.1 主要工作总结

本论文将个性化推荐技术引入到网络求职招聘中，能提高求职者获取信息的即时性和投递简历的准确性。主要研究内容包含以下几个方面：

#### (1) 求职者过程模型的研究

以 Soelberg 和 Kanfer 定义为理论依据，将求职定义为一个不断自我调试的过程，并提出以求职者的职位竞争力指数和职位适合度指数为个人自我评价和推荐的依据。

#### (2) 求职者求职意愿模型的研究

采用实证分析和聚类方法来构建求职者求职意愿模型。通过对中华英才网、智联招聘网、前程无忧网等国内知名求职网站的求职意愿属性进行调查，综合得到求职意愿属性的全集。

#### (3) 求职者职位竞争力评价模型的研究

通过分析职位竞争力的源起、目标及意义确立了以职位信息为核心，综合考虑个人基础信息、教育背景、获奖情况、语言能力的 4+1 评价指标模型，并采用基于优势关系的粗集多属性排序方法评判评价指标，最终得到求职者的职位竞争力。

#### (4) 求职者职位适合度评价模型的研究

通过分析问题特征发现该问题与职位竞争力模型属于同一类问题，因此采用与求职者竞争力模型相同的方法建立该模型。

#### (5) 招聘信息个性化推荐系统设计与实现

设计并实现了网络职位信息个性化推荐系统，其中智能体 Agent 以“持续查询”策略不断监测求职者简历和职位的信息变化，并计算求职者的竞争力和职位适合度，最后根据求职者的兴趣爱好等，按照不同策略为求职者推荐不同的信息。

### 5.2 心得体会

通过本次毕业设计，我深深感受到了科学研究过程中需要的一丝不苟、严谨求实、精益求精的态度。首先在论文研究初期，学会了如何在大量的资源中选择

和查阅文献、写综述；在论文研究过程中，遇到问题、挫折时要保持积极面对的心态去解决问题；在论文的写作过程中，学会了如何既突出重点，又能精确、简洁、注意文法规范的写作方法。

### 5.3 进一步开展工作

#### (1) 引入机器学习的方法

期望在后绪研究中将机器学习引入到求职者求职意愿模型，以便能根据求职者的浏览记录自动获取求职者的求职意愿，进一步提高推荐的准确性。

#### (2) 双向推荐问题

本文主要是向网络求职者推荐职位信息，因此下一步可以考虑向网络招聘者推荐求职者的简历信息，以便为企业推荐更适合的人才。



## 参考文献

- [1] Steffy B D, Shaw K N, Noe A W. Antecedents and consequences of job-search behaviors. *Journal of Vocational Behavior*, 1989, 35(3): 254-269
- [2] Bretz R D, Boudreau J W, Judge T A. Job-search behavior of employed managers. *Personnel Psychology*, 1994, 47(2): 275-301
- [3] Soelberg P O. Unprogrammed Decision Making. *Industrial Management Review*, 1967, 8(2): 19-29
- [4] Kanfer R, Wanberg C R, Kantrowitz T M. Job search and employment: A personality-motivational analysis and meta-analytic review. *Journal of Applied Psychology*, 2001, 86(5): 837-855
- [5] 沈士仓, 姜澎. 我国网络招聘的现状与问题. *中国人才*, 2002.
- [6] 曾春, 邢春晓, 周立柱. 个性化服务技术综述. *软件学报*. 2002.
- [7] 袁庆宏, 刘昌苹. 国内外网络招聘行业现状分析. *信息技术*. 2009.
- [8] 第 25 次中国互联网络发展状况统计报告. 2010.
- [9] 吴丽花, 刘鲁. 个性化推荐系统用户建模技术综述[J]: *情境学报*, 2006(1): 55-62.
- [10] Jokela, S Turpeinen. M The role of structured content in a personalized news service *Proceedings of the Hawaii International Conference on System Sciences*, 2001.
- [11] Lee. H-R Hierarchical filtering method for content-based music retrieval via acoustic input *Proceedings of the ACM International Multimedia Conference and Exhibition*, 2001.
- [12] Chedrawy, zeina. A Web recommender system for recommending, predicting and personalizing music playlists 10th International Conference on Web Information Systems Engineering, 2009.
- [13] Linden, Greg Smith, Brent Amazon. com Recommendations: Item-to-item collaborative filtering *IEEE Distributed Systems Online*, 2003.
- [14] Mudambi, S. M. What makes a helpful online review? A study of customer reviews on amazon. com *MIS Quarterly: Management Information Systems*, 2010.
- [15] 严慧英. 影响网络信息检索行为的主体因素. *情报杂志*, 2004.
- [16] Lieberman, H. Letizia. An agent that assists web browsing *Proceedings of the International Joint Conference on Artificial Intelligence*, 1995.
- [17] 林霜梅, 汪更生, 陈弈秋. 个性化推荐系统中的用户建模及特征选择. *计算机工程*, 2007 (9): 196-198.

- [18]姜有辉,高琳琦. 个性化推荐系统中顾客信息的隐式采集方法研究. 现代情报, 2006 (11):10-13.
- [19]Makris,Christos. An integrated Web System to facilitate personalized web searching algorithms 23rd Annual ACM Symposium on Applied Computing, 2008.
- [20]陈琳娜. 基于本体的个性化用户模型研究. 燕山大学, 2006.
- [21]Fong,Simon. On designing a Market monitoring Web Agent system Proceedings of the 10th International Conference on Information Integration and Web-based Applications and Services ,2008.
- [22]Pazzani. Michael Syskill & Webert: identifying interesting web sites .Proceedings of the National Conference on Artificial Intelligence,1996.
- [23]朱祥玉. 基于向量空间模型的自适应文本过滤系统研究. 山东师范大学, 2006.
- [24]宋炜,张铭. 语义网简明教程. 北京:高等教育出版社, 2004.
- [25]韦鲁玉. 基于 Agent 的个性化智能信息检索系统. 哈尔滨理工大学, 2007.
- [26]代学武. 基于神经网络的用户建模和 Web 信息过滤研究. 西南师范大学, 2003.
- [27]Wu, Y.H., Chen, Y.C., Chen, A.L.P. Enabling personalized recommendation on the web based on user interests and behaviors. In: Klas, W., ed. Proceedings of the 11th International Workshop on Research Issues in Data Engineering. Los Alamitos, CA: IEEE CS Press, 2001.
- [28]周明,孙树栋. 遗传算法原理及应用. 北京: 国防工业出版社, 2002.
- [29]DMladenic Personal Web Watcher:Design and Implementation Technical Report ,1996
- [30]O. de Vel, S. Nesbitt .A Collaborative Filtering Agent System for Dynamic Virtual Communities on the Web Conference on Automated Learning and Discovery, 2003
- [31]Resnick P, Iakovou N, Sushak Met. GroupLens: An open architecture for collaborative filtering of netnews. Proc 1994 Computer Supported Cooperative Work Conf, Chapel Hill ,1994.
- [32]P Cunningham, WEBSSELL : Intelligent Sales Assistants for the World Wide Web Technical Report, 2001.
- [33]Will Hill, Larry Stead. Recommending and Evaluating Choices in a Virtual Community of Use Conference on Human Factors in Computing Systems archive, 1995.
- [34]Miller, Bradley N MovieLens unplugged: Experiences with an occasionally connected recommender system International Conference on Intelligent User Interfaces, Proceedings, 2003.
- [35]单松巍, 冯是聪, 李晓明. 几种典型特征选取方法在中文网页分类上的效果比较. 计算机工

程与应用, 2003.

[36]赵鹏, 耿焕同, 王清毅, 蔡庆生. 基于聚类和分类的个性化文章自动推荐系统的研究. 南京大学学报. 2006, 9 (5) :512-518.

[37]张培颖. 基于 Web 内容和日志挖掘的个性化网页推荐系统. 计算机系统学报, 2008 (9) :9-11.

[38]王虹予, 赵英, 党跃武. 基于混合算法的电子商务推荐系统设计研究[J]. 知识组织与知识管理, 2009 (1) :80-85.

[39]崔林, 宁瀚涛, 陆玉昌. 基于语义相似性的资源协同过滤技术研究. 北京理工大学学报. 2005 (5) :402-405.

[40]王宏宇, 糜仲春, 梁晓艳, 叶跃祥. 一种基于支持向量机回归的推荐算法. 中国科学院研究生院学报, 2007.

[41]冀俊忠, 沙志强, 刘椿年. 贝叶斯网模型在推荐系统中的应用研究. 计算机工程, 2005 (7) :32-34.

[42]Somlo G, Howe A. Adaptive lightweight text filtering. Proc Lecture Notes In Computer Science, 2001: 319-329.

[43]刘建国, 周涛, 汪秉宏. 个性化推荐系统的研究进展. 自然科学进展, 2009.

[44]徐泽水. 求解不确定型多属性决策问题的一种新方法. 系统工程学报, 2002.

[45]徐玖平, 吴巍. 多属性决策的理论与方法, 北京: 清华大学出版社, 2006.

[46]戴云龙, 王俊生. 基于层次分析法的多因素综合选课模型. 软件开发, 2009.

[47]葛慧明. 层次分析法在专业技能评估中的应用. 无锡南洋学院学报, 2008.

[48]徐志新, 奚树人, 曲静原. 核事故应急决策的多属性效用分析方法. 清华大学学报, 2008.

[49]罗万成, 吴朝平. PC-LINMAP 模糊综合评判学生工作绩效评估模型. 重庆文理学院学报, 2006.

[50]李宁霞, 谢定华, 董鹏. LINMAP 方法在生产企业供应商选择中的应用. 中国管理科学, 2009.

[51]钱挹清. 应用模糊综合评判法进行东莞市水资源规划宏观经济社会效益评价. 珠江现代建设, 2006.

[52]刘学生. 基于粗集的不确定多属性决策排序法的研究. 大连理工大学, 2009.

[53]史战红, 连玉平. 基于包含度的粗糙集间的相似性度量. 数学教学研究, 2008.

[54]路艳丽, 雷英杰. 直觉模糊集的包含度. 计算机科学, 2009.

## 致谢

经过一年多的学习与研究，论文终于得以完成。首先要感谢我的论文指导教师林正奎老师对我的耐心指导和帮助。不仅从论文选题、资料的收集、论文的结构到论文的最终定稿，林老师都给了我悉心的指导、中肯的建议和修改的意见。

同时，还要感谢我的导师孟丽艳老师在生活、学习等各方面给予的无微不至的关怀和帮助。

还有我的师兄肖智博在我论文研究过程中也给予了很大的帮助，特别是论文系统实现阶段。好友黄会、贺鑫等也给我提出了好多建议，张安娜师妹、易忠伟师弟等在我论文的写作过程中给予了很大的帮助，在此也向他们表示衷心的感谢。

最后，我还要感谢我的父母对我完成论文的督促与指导，以及对我平时学习的关心。

# 网络招聘信息个性化推荐技术研究

作者：[张维瑞](#)  
学位授予单位：[大连海事大学](#)

## 本文读者也读过(2条)

1. [邓一飞](#) [中国网络招聘行业现状研究及发展分析](#)[学位论文]2006
2. [朱蓉蓉](#) [求职者对网络招聘态度及其影响因素研究](#)[学位论文]2008

本文链接：[http://d.wanfangdata.com.cn/Thesis\\_Y1696730.aspx](http://d.wanfangdata.com.cn/Thesis_Y1696730.aspx)