

学校编码: 10384

分类号 _____ 密级 _____

学号: 23020131153210

UDC _____

厦 门 大 学

硕 士 学 位 论 文

基于模糊理论的人才招聘系统研究以及关联规则在其中的应用

Research of Recruitment System Based on Fuzzy Theory
and the Application of Association Rules

应 文 涛

指导教师姓名: 吴梅红 副教授

专业名称: 计算机技术

论文提交日期: 2016 年 04 月

论文答辩时间: 2016 年 04 月

学位授予日期: 2016 年 月

答辩委员会主席: _____

评 阅 人: _____

2016 年 04 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文,并向主管部门或其指定机构送交学位论文(包括纸质版和电子版),允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索,将学位论文的标题和摘要汇编出版,采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于:

() 1. 经厦门大学保密委员会审查核定的保密学位论文,于 年 月 日解密,解密后适用上述授权。

() 2. 不保密,适用上述授权。

(请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文,未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的,默认为公开学位论文,均适用上述授权。)

声明人(签名):

年 月 日

摘 要

无论是小到各行各业，还是上升到国际的高度，所有的竞争，归根结底，实质均是人才的竞争，传统的人才招募方式已经不能满足时代的需要，会逐步被社会所淘汰；新的十三五规划纲要提出，我国要发展当代互联网产业结构，贯彻落实“互联网+”战略，更深更广地使用互联网；目前已有的若干招聘网站尚有一些可改进的地方；如何从繁多的人才简历资料中发现其中各项数值彼此的关联关系，并归纳出可利用的信息来服务双方，是一个迫在眉睫的问题。

结合以上这几点，本文通过研究以模糊集理论为基础的简历-职位匹配系统，同时从两个方向优化了 Apriori 算法，最终设计并实现一个以互联网为载体，B/S 模式的人才招聘系统，其主要工作如下：

(1) 首先，结合国内外网络招聘的现状与数据挖掘技术的应用情况，讨论了人才招聘系统的研究价值与意义，并说明了几个实现所需的主要平台技术。

(2) 其次，阐述了关联规则的概念与基础，讨论了 Apriori 算法与 FP-growth 算法的适用范围及优劣，在两处地方对 Apriori 算法实行了优化，并介绍了模糊集合论的相关知识理论与理论。

(3) 然后，对系统实施了的多方面的分析，并依据分析的思路进行了系统整体的设计、数据库的设计等；根据对模糊产生型规则的分析与数据特征，使用一种加权的匹配算法设计了简历-职位匹配系统。

(4) 完成系统的实现，举例阐明了简历与职位之间的匹配过程，并把本文阐述的经过优化的 Apriori 算法应用在收集到的简历资料中，得到若干能够供双方参考的信息。

简历-职位匹配系统可以为求职者推荐适合的职位，大幅削减了求职者的时间成本。最后把本文阐述的经过优化的 Apriori 算法与传统的 Apriori 算法的效率做了对比，对比结果显示优化后的 Apriori 算法更有价值。

关键词：人才招聘；模糊集理论；Apriori 算法

Abstract

Whether it is small to all trades and professions, or up to the international level, the essence of all the competition is the competition of talents in the final analysis, The traditional recruitment methods cannot meet the needs of the times, it will be eliminated by the society gradually; The 13th Five-year plan proposed to develop a modern industrial system, the implementation of the “Internet plus” action plan to promote the wide application of the Internet; There are still something can be improved in existing recruitment websites; It is an urgent problem how to find out the relationship between each item of so many resume data, and summarize available information to serve both sides.

Considering the points above, this dissertation designed and implemented an Internet-carrier recruitment system with B/S mode by studying fuzzy set theory based cv-job matching system, and the Apriori algorithm is optimized from two directions at the same time. The main work is as follows:

(1) First of all, according to the status of online recruitment and application of data mining technology, discusses the value and significance of the topic and explains several major technologies required to achieve the system.

(2) Secondly, the concept of association rules is elaborated and discussed the applicable scope, pros and cons of Apriori algorithm and FP-growth algorithm, improve the Apriori algorithm in two places, and introduces the relevant knowledge and theory of fuzzy set theory.

(3) Thereafter, the system implemented a multifaceted analysis, and the overall design, database design based on the analysis of ideas. According to the analysis of fuzzy production rules and the data feature, a weighted matching algorithm is used to design cv-job matching system.

(4) The realization of the system is completed, we illustrate the matching process between resume and job. Apply the approved Apriori algorithm which we illustrates to the resume information collected by a company, we obtain a number of information

what can be used for both sides.

Cv - job matching system can recommend appropriate jobs for job-seekers, a significant reduction in time cost of job seekers. And we compare the improved Apriori algorithm with the traditional Apriori algorithm, comparison results show that the improved Apriori algorithm is more valuable.

Keywords: Online Recruitment; Fuzzy Sets Theory; Apriori Algorithm

厦门大学博硕士学位论文摘要库

目 录

第一章 绪论	1
1.1 研究背景与意义	1
1.2 国内外研究现状	2
1.2.1 网络招聘的研究现状.....	2
1.2.2 数据挖掘技术的研究现状.....	4
1.3 本文主要工作.....	5
1.4 本文结构安排.....	5
第二章 系统相关技术介绍	7
2.1 Visual Studio 2010	7
2.2 SQL Server 2008 R2 数据库技术.....	8
2.3 ASP.NET 技术.....	9
2.4 ADO.NET 技术.....	10
2.5 B/S 模式.....	11
2.6 本章小结	12
第三章 关联规则算法与模糊理论集	14
3.1 关联规则技术概述.....	14
3.1.1 关联规则的基本概念.....	14
3.1.2 关联规则的一般步骤.....	15
3.1.3 关联规则类别.....	16
3.2 关联规则经典算法介绍.....	17
3.2.1 Apriori 算法	17
3.2.2 FP-growth 算法	21
3.3 Apriori 算法的改进.....	23
3.4 模糊集合论.....	26
3.4.1 模糊集定义.....	27
3.4.2 模糊集理论基础.....	27

3.5 本章小结	29
第四章 系统需求分析与设计	30
4.1 系统概述	30
4.2 需求分析	30
4.2.1 功能需求分析	30
4.2.2 业务需求分析	32
4.2.3 其他类型需求分析	37
4.3 系统设计的目标和原则	38
4.4 系统整体设计	38
4.4.1 系统架构设计	38
4.4.2 功能模块设计	39
4.5 数据库设计	40
4.5.1 数据库设计原则	40
4.5.2 数据库概念构造设计	41
4.5.3 数据库逻辑构造设计	44
4.6 基于模糊理论的职位推荐设计	48
4.6.1 加权模糊集合论介绍	48
4.6.2 简历-职位匹配模块设计	50
4.7 本章小结	54
第五章 系统的实现	55
5.1 功能实现与界面设计	55
5.2 基于模糊理论的职位推荐	60
5.3 Apriori 算法在招聘中的应用	64
5.3.1 数据预处理	64
5.3.2 结果与分析	66
5.4 本章小结	68
第六章 总结与展望	69
6.1 总结	69

6.2 展望.....	69
参考文献.....	71
致 谢	76

厦门大学博硕士论文摘要库

Contents	
Chapter 1 Introduction	1
1.1 Background and Significance	1
1.2 Current Research.....	2
1.2.1 Status of Online Recruitment	2
1.2.2 Status of Data Mining	4
1.3 Main Work	5
1.4 Structure	5
Chapter 2 Related Technologies	7
2.1 Visual Studio 2010	7
2.2 SQL Server 2008 R2	8
2.3 ASP.NET.....	9
2.4 ADO.NET.....	10
2.5 B/S.....	11
2.6 Summary	12
Chapter 3 Association Rules and Fuzzy Set Theory.....	14
3.1 Introduction of Association Rules.....	14
3.1.1 Concept of Association Rules	14
3.1.2 Procedure of Association Rules	15
3.1.3 Types of Association Rules	16
3.2 Introduction of Classical Algorithms	17
3.2.1 Apriori Algorithm.....	17
3.2.2 FP-growth Algorithm	21
3.3 Improved Apriori Algorithm.....	23
3.4 Fuzzy Set Theory	26
3.4.1 Definition of Fuzzy Set.....	27
3.4.2 Theoretical Principle of Fuzzy Set.....	27

3.5 Summary	29
Chapter 4 Requirements Analysis and Design	30
4.1 System	30
4.2 Requirements Analysis	30
4.2.1 Functional Requirements	30
4.2.2 Business Requirements	32
4.2.3 Other Types of Requirements	37
4.3 System Design Goal and Principles	38
4.4 System Overall Design	38
4.4.1 Structure Design	38
4.4.2 Functional Design	39
4.5 Database Design	40
4.5.1 Design Principles	40
4.5.2 Conceptual Database Design	41
4.5.3 Logical Database Design	44
4.6 Position Recommendation Design Based on Fuzzy Set Theory	48
4.6.1 Introduction of Weighted Fuzzy Set Theory	48
4.6.2 CV-Job Matching Module Design	50
4.7 Summary	54
Chapter 5 System Implementation	55
5.1 Functional Implementation and Interface Design	55
5.2 Position Recommendation Based on Fuzzy Theory	60
5.3 Application of Apriori Algorithm In Recruitment	64
5.3.1 Data Preprocessing	64
5.3.2 Results and Analysis	66
5.4 Summary	68
Chapter 6 Conclusions and Outlook	69
6.1 Conclusions	69
6.2 Outlook	69

References 71

Acknowledgment..... 76

厦门大学博硕士论文摘要库

第一章 绪论

1.1 研究背景与意义

大步踏入二十一世纪之后，整个社会的物质、文化的成长均登上了一个更巍峨的台阶，愈来愈多的普通人把握住并取得了职业教育或高等教育的机会，通过研习掌握了一门或数门工作的才略。不容置喙，这抬升了全社会的素质水准与职业层次，对推进我们社会整体的变更有重大利好，但不可否认的是，伴随着技工和大学生数额的井喷，他们在就业方面的焦虑与日俱升。为了一个理想的职位而奔走于全国各省，耗费了繁多的时间、钱财和精力，最终还可能一无所获，如此的同学朋友并不鲜见。甚至有些人还停留在从纸质媒体上搜罗招聘广告信息的层面，一份报纸上印刷的招聘相当有限，这样做对求职者而言费力费时，对公司来说不仅需要支付不菲的版面广告费，收到的反馈还特别有限，招聘效率极其低下，大规模的油墨印刷也对环境不利。

互联网作为科技成长的果实，现已融入到生活中的各个角落。除了一些特定需要超级保密的文件为了防止被黑客窃取而不会备份在互联网上以外，大多数文件数据已经做到信息化数字化了。很多大型公司与互联网企业已经逐步地运用网路进行办公和人才招聘^[1]。招聘网站既是面向求职者，也是面向求贤者们。一方面，需求方可以借由该平台及时发布公司最新的求贤信息，操作简便，不会被地点、时间所绑架，并且在较高程度上节俭了用人方的资源，锐减了公司的成本花销，增进了公司的办事效率^[2]。一次考虑周到，近乎完美的网络招聘还能提升公司在职场人心中的形象。另一面，求职者们不需再受车马奔波之劳，不用再为错过某场心仪的公司招聘会而捶胸顿足，只需在有网络的设备上浏览琳琅满目的招聘咨询，进行个人注册，建立电子简历并投递到所向往的公司，然后等待后续进度并根据自己的计划进行安排即可。

之前《财富》报刊举办了一个统计，国际 500 强的集团内的岗位里，有 88% 的岗位是通过网络来招聘的^[3]。网上招聘的便捷、高效、省力、无纸化等特征让人们尝到甜头后，人们更会倾向于在以后的工作中采用互联网来实施应聘与招聘，

通过报纸、杂志等媒介的招聘方式会渐行渐远。随着在网路上投递简历的人愈来愈多，招聘网站的数据仓库里会堆积成千上万的电子简历，有些甚至从未被打开过。平时我们只能看到这海量简历数据呈现出的简单的表面特征，其身后隐匿的大批待利用的知识，迫切需要人们去寻找与剖析，用来更好地支持公司的招聘决策。知识发现与数据挖掘（Knowledge Discovery and Data Mining）这一技术恰好就是用来处理这一类问题的，已然成为最近几年关注度较高的一个研究热点^[4]。其中，发掘相关联的规则是数据挖掘领域的一项十分关键的分支^[5]。

本文研究的中心意义在于转变传统概念里的招聘与求职行径，利用网络高效、快捷的优势，打破空间与时间的制约，为双方搭建起沟通交流的网上桥梁，并深层次地学习数据挖掘技术里典型的关联规则算法，将改良过的 Apriori 算法运用到广阔的简历数据中，探索隐藏在简历冰山下的信息，以便公司及时地调整自身的招聘策略，洞悉市场局势与偏好，进行更好更佳的人才选拔。

1.2 国内外研究现状

1.2.1 网络招聘的研究现状

从此前智联招聘发布的数据来看，从 2005 年到 2012 年，我国线上招聘占整个招聘市场的规模比重一直在提高，在 2012 年到达 19.6%，当年市场体量是 180 亿元^[6]。经过数年的竞争和发展，我国市场目前有几个大型成熟的专业招聘网站，例如 51job、中华英才网等等^[7]，它们都拥有专业的线上招聘系统，专精于做线上招聘，提供求职和人才招聘服务。还有一些综合性较强的站点例如 58 同城和赶集网，也有提供网络招聘的功能，只是并不是专业的服务，只能适用于地方性的小企业的全职或兼职招聘。还有一部分大型企业由于招聘需求大而且是全国范围内进行招聘，纷纷设立了自己的专用招聘网站，例如中国农业银行、中国移动等等^[8]。这些系统和网站推进了我国的人才招聘，加速了人才流动，特别是在我国高等院校的校招过程中，线上招聘已然成为主流。

根据腾讯发布的《2015 年中国网络招聘市场调查》显示，目前用户得到工作的方式特别多，其中最有效的方式是通过招聘网站，占比 31.34%^[9]。而从用户获得招聘信息的渠道来源来看，专门的招聘网站以 53.71% 的占比的绝对优势成为

第一，超过第二的“亲朋好友介绍”将近 21 个百分点，可见网络招聘已经成为普通人最受欢迎、最频繁使用的求职渠道。

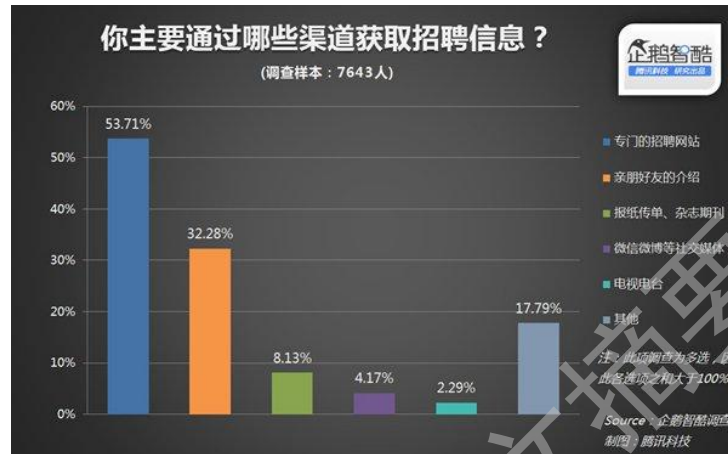


图 1.1 2015 年中国网络招聘渠道调查

在这么多种类的应聘方式里，在线招聘系统的份额颇为杰出，但是人们对它的负面反馈也达到最多，在线招聘系统存在已久的弊端还是有待解决，例如一直以来的虚假广告、沟通过程缓慢、使用体验差以及界面设计落后等不足。从图中可以看出，虚假广告的投诉率最高，反馈信息太少与沟通冗长的比重也很高——当今在线招聘系统要走的路还很长。

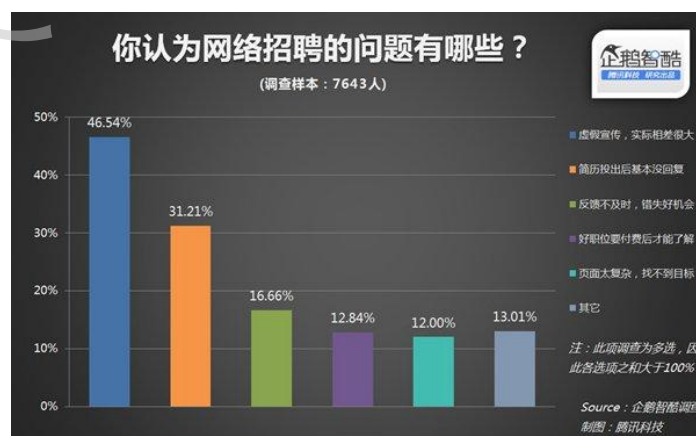


图 1.2 2015 年中国网络招聘问题调查

在一些发达国家里，网络人才的招聘发展得尤其迅速和先进。按照网络信息

中心的报告, 2001 年, 通过网络雇用员工的市场体量只有 27 亿美元, 2004 年上升到 134 亿美元, 2008 年增加到 198 亿美元, 在 2011 年的体量超过了 450 亿美元^[10], 增长速度令人咋舌。全球知名的招聘网站, 例如 Monster.com 和 linkedin.com, 前者是全球范围内最大的网络求职站点, 根据 comScore 媒体 2006 年的报告, 它是现今世上 20 个点击率最高的站点之一; 而后者则是世界上最大的职业社交网站, 除了招聘求职的功能外, 还附加了社交的功能, 可以寻找同学、同事、合作伙伴, 建立并拓展人脉网络, 开发自己的职业潜力, 国内译名领英^[11]。甚至还有公司已经研发出自己的人才测评系统, 应聘者只要完整地参加整套网络测试, 系统就能判断出该应聘者是否是公司想要的人才, 而不再需要额外的面试过程。可见, 发达国家的招聘工作已经主要在网络上进行, 而不再是传统的线下人工招聘。

1.2.2 数据挖掘技术的研究现状

在国外, 对 KDD 的研究主要环绕着技术、理论和应用这三大方向铺开^[12]。多种理论和方式的结合是绝大多数研究人员选用的有效的技术。目前最新的发展主要是对探索知识的方式的更深层次的研究, 比如近年来重视对 Bayes 方法和 Boosting 方法的研究和改良; KDD 和 Database 的密切结合; 古老的统计学的回归方式在 KDD 里的应用。而在应用的一方主要表现在 KDD 商业软件从搞定问题的独立过程转为建设解决问题的一体化系统, 重点用户有大型银行、保险企业和销售公司等。SAS 公司研制的 Enterprise Miner 是在数据挖掘市场上首屈一指的竞争者, 是一个及其强大的开采信息的产品, 包含群、关联、Decision Tree、network of neuron 与传统的 RA 技术, 简单易用, 在市场上赢得了硕大的成功^{[13][14]}。大量的计算机企业和研究机构均十分重视数据挖掘的学习和应用的重要性, 例如微软就建立了自己的研究中心。

相对于发达国家, 中国对于数据挖掘技术的摸索起步稍慢一些, 尚且不够成熟, 正处于一个发展的过程。国内的许许多多院校实验室、国家级的科研单位和企业研究所都在学习数据挖掘的基本理论。比较新的发展成果有: 使用概念来实现文本挖掘; 探讨中文文本的挖掘模型理论和相应的实现技术; 建立智能化的研究系统; 建立模糊系统识别方法和模糊系统的理论知识模型; 把粗糙集以及模糊

Degree papers are in the “[Xiamen University Electronic Theses and Dissertations Database](#)”.

Fulltexts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.