

·专题: 信息管理与电子商务·

# 基于大数据的需求驱动的职业能力培养研究



□刘 耘<sup>1</sup> 袁 华<sup>2</sup>

[1. 四川交通职业技术学院 成都 611130; 2. 电子科技大学 成都 611731]

**[摘 要]** 提出了一个基于大数据的人才培养方案, 该方法首先从互联网上海量的招聘信息中获取该类别岗位的能力需求特征; 接着在制订人才培养方案时遵循岗位类别——能力需求特征——知识结构、能力结构和素质结构的路径来设计培养目标和培养规格, 从而实现需求驱动的职业能力的培养。

**[关键词]** 大数据; 主题模型; PLDA; 需求驱动的职业能力培养

**[中图分类号]** G645

**[文献标识码]** A

**[DOI]** 10.14071/j.1008-8105(2018)-1001

## Research of Demand-Driven Vocational Capacity-Building Based on Big Data

LIU Yun<sup>1</sup> YUAN Hua<sup>2</sup>

(1. Sichuan Vocational and Technical College of Communications Chengdu 611130 China;

2. University of Electronic Science and Technology of China Chengdu 611731 China)

**Abstract** A method based on big data to achieve demand-driven talents cultivation is presented. Firstly, skill requirements are obtained from the recruitment information on the Internet. Secondly, a roadmap is defined to design the training objective and specification, which is from position to skill requirements and then to knowledge structure, ability structure, and quality structure, so as to realize the demand-driven vocational capacity-building.

**Key words** big data; topic model; PLDA; demand-driven vocational capacity-building

领英(LinkedIn)是全球最大的职业社交网站, 其创始人及产品管理副总裁Allen Blue认为, 创新正在以前所未有的速度在全球范围内展开, 各国的教育体系都在加紧脚步以适应这种巨变, 以弥补在所有产业中员工自身的职业能力与企业所需之间的差距<sup>[1]</sup>。

2014年教育部等六部门组织编制了《现代职业教育体系建设规划(2014~2020年)》, 该规划指出职业教育应坚持以服务发展为宗旨, 以促进就业为导向<sup>[2]</sup>。2015年修订的《普通高等学校高等职业教育(专科)专业目录》在进行专业划分和调整时, 原则上专业大类对应产业, 专业类对应行业, 专

业对应岗位群<sup>[3]</sup>。然而职业技术学院在制订专业的人才培养方案时, 要比较准确地把握某一类别岗位的能力需求特征又存在一定的难度, 传统上都是使用问卷调查和专家访谈等方法。这些方法存在样本小、费时费力等缺点, 并且调研结果往往具有较大的局限性。

世界银行旗下的研究机构Education Global Practice Group在其《走向劳动力市场政策2.0》一文中提出了利用大数据来实现“需求驱动的职业能力培养”的思想, 该文作者Nomura等通过对互联网上的招聘信息进行文本分析, 更加深入地了解企业对员工的职业能力的需求<sup>[4]</sup>。后续的研究与之类

[收稿日期] 2018-01-15

[基金项目] 四川省教育厅2017年度人文社科重点项目(17SA0112)。

[作者简介] 刘耘(1970-)女, 博士, 四川交通职业技术学院信息工程系副教授; 袁华(1973-)男, 博士, 电子科技大学经济与管理学院副教授。

似,都把重点放在如何利用大数据获取需求上,鲜有进一步探讨需求怎样驱动职业能力的培养。

本文将这种基于大数据的方法用于辅助职业技术学院制订专业的人才培养方案,帮助他们“主动适应、服务发展”<sup>[2]</sup>,特别是当这些专业面向的是新兴或快速发展的产业时。本文的方法由两部分组成,第一部分是从互联网上海量的招聘信息中获取某一类别岗位的能力需求特征;第二部分是在制订人才培养方案时遵循岗位类别——能力需求特征——知识结构、能力结构和素质结构的路径来设计培养目标 and 培养规格,从而实现需求驱动的职业能力的培养。

本文首先对相关研究进行文献回顾,然后详细介绍本文提出的方法,接下来分析软件技术专业所针对的岗位类别(软件工程师)的能力需求特征,并且对四川交通职业技术学院制订的人才培养方案进行修订,结果表明我们的方法具有相当程度的实用性,最后,进行总结和展望。

## 一、文献回顾

与在报纸上刊登招聘广告不同,企业在招聘网站发布信息不需要按字付费,这样就能够更详细地反映它们要求的技能<sup>[5]</sup>。互联网上的招聘信息不仅样本量大,还具有实时性和动态性,因此能够用来分析那些传统数据无法分析的问题<sup>[6]</sup>。数据的获取也更加灵活、快速、简单和低廉<sup>[7]</sup>。

Chan等从1999年5月~2000年1月在australianjobsearch.com上抓取与电子商务有关的招聘信息,通过描述性统计分析阐述了不同岗位类别的职责和需要掌握的技能<sup>[8]</sup>。Wade等同时在两家贸易杂志和五家招聘网站上共抓取了800个招聘信息,其岗位类别为网站管理员。采用内容分析法得到的结果显示企业对该岗位类别的技能要求是多层面的,包括专业技能和组织技能,而且对专业技能的要求高于对组织技能的要求<sup>[9]</sup>。这两个研究比较类似,都是采用人工的方法对招聘信息的文本进行分析。

Capiluppi等认为,企业希望员工具备的不是笼统的IT技能,而是像“Java、.NET、C++”这样具体的技能。他们从monster.com上抓取了48000个与IT有关的招聘信息,接着从招聘信息中抽取出岗位需求,然后使用一个专业术语词典和一系列正则表达式识别出30个最常出现的词<sup>[10]</sup>。

Wowczko着重描述了利用大数据进行分析的步骤。他从2014年1月~2014年12月在irishjobs.ie上抓

取了7090个与IT有关的招聘信息。在经过一系列数据预处理后分别得到了岗位名称和岗位需求两个语料库,接下来在文本分析时首先把岗位名称语料库中最常出现的词挑选出来作为岗位类别,然后在岗位需求语料库中增加一个标签列,该列的值来自于岗位名称语料库,但是根据岗位类别进行了合并和去重。就这样确定了岗位需求语料库中4755个的类别,接下来利用K-最近邻分类算法对剩下的2228个进行分类。最后得到了7个不同类别岗位的前20个常用技能。Wowczko认为,只使用岗位名称存在不少局限,好的方法是同时利用岗位需求中丰富的信息<sup>[11]</sup>。

为了理解大数据的技能需求,Mauro等从dice.com上抓取了包含“big data”的2786个招聘信息。在进行文本分析时,首先把岗位名称语料库中相邻的两个词(Bigram)按照频次排序,再由专家评选,确定4个岗位群(例如数据科学家岗位群包括了数据科学家、数据工程师和数据分析师3个岗位)。然后建立隐含狄利克雷分布(Latent Dirichlet Allocation, LDA)模型对岗位需求语料库进行聚类,得到了9个技能组(例如项目管理技能组包括了风险管理、计划等技能)。最后通过模型分析岗位群与技能组的相关性,揭示出9个技能组在每个岗位群中的重要程度<sup>[12]</sup>。

Nomura等认为企业很少直接询问求职者有关非认知技能的问题,因为缺乏衡量这些技能的客观标准。但是在对专家进行访谈或做问卷调查时,可以了解到他们存在这种需求。而分析招聘信息可以得到对需求的更准确的描述<sup>[4]</sup>。

由于缺乏数据,国内的研究还处于探索阶段。程茜等从校园网获取招聘信息,由于招聘信息是以文本的形式发布,所以他们采用人工的方法对其进行分析<sup>[13]</sup>。钟晓旭等使用网络爬虫从“新安人才网”抓取招聘信息,但他们只对岗位类别、学历等属性进行了分析<sup>[14]</sup>。只有彭金莲等对招聘信息中的岗位需求进行了文本分析,但从中只识别出40个与网络工程有关的术语<sup>[15]</sup>。

## 二、研究方法

### (一) 获取岗位的能力需求特征

要获取岗位的能力需求特征,关键是对招聘信息进行文本分析。有关步骤如图1所示。

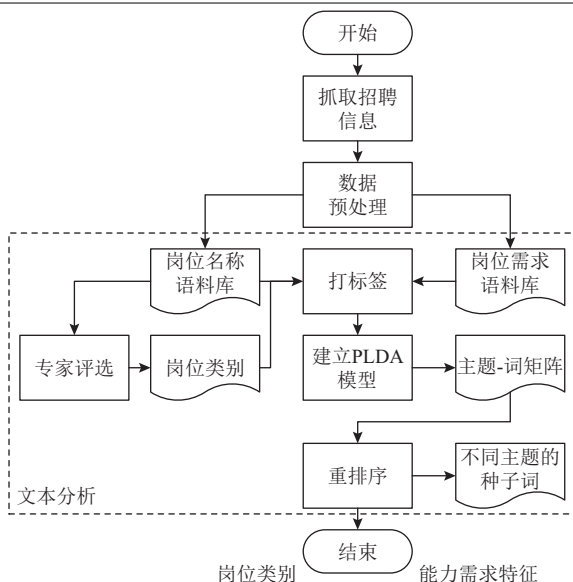


图1 获取岗位的能力需求特征的步骤

### 1. 抓取招聘信息

为了从互联网上抓取招聘信息,我们开发了网络爬虫系统。这个系统可以定时到招聘网站下载网页,然后对网页进行解析,从中提取出招聘信息。

一条招聘信息中包含了岗位名称、岗位类别、公司名称、公司性质、公司规模、所属行业、招聘人数、发布日期、岗位职责和岗位需求等属性,其中的岗位名称和岗位需求都是文本,需要进行下面的数据预处理。

### 2. 数据预处理

#### (1) 中文分词

在进行文本分析前,需要进行中文分词。本文选择的是IKAnalyzer,在分词的同时进行了大小写转换。

#### (2) 拼写检查

岗位需求中常常会出现拼写错误的词,本文采用编辑距离算法来发现这些错误的词。

#### (3) 同义词替换

在纠正了拼写错误后,还要进行同义词替换可以降低数据的维度。

经过以上预处理,就得到了岗位名称语料库和岗位需求语料库。这两个语料库分别有两列,一列是ID,代表招聘信息;另一列是岗位名称或岗位需求在预处理后的文本。

### 3. 文本分析

#### (1) 打标签

把岗位名称语料库中的所有词按照频次排序,再由专家评选,确定岗位类别。接下来在岗位需求语料库中增加一个标签列,该列的值来自于岗位名称语料库,但是根据岗位类别进行了合并和去重。标签列的值可能有多个,也可能没有。

### (2) 建立PLDA模型

招聘网站上的招聘信息都有岗位类别这个属性,但是不够精确。例如,岗位类别为软件工程师的招聘信息,其岗位名称可以具体为Java软件工程师。所以,我们认为岗位类别应该进一步细分。

在上面打标签之后,岗位需求语料库中保存的是带标签的预处理后的文本,这样就可以采用有监督学习或半监督学习分类算法来得到不同类别的岗位对应的能力需求特征。由于一个招聘信息可能同时针对多个类别的岗位,所以这是一个多标签分类问题。而主题模型(Topic Model)就是一种多标签分类的概率生成模型。

主题模型的起源是隐性语义分析(Latent Semantic Analyzing, LSA)/隐性语义索引(Latent Semantic Indexing, LSI)。在LSA/LSI的基础上,Hofmann提出了概率隐性语义索引(probabilistic Latent Semantic Indexing, pLSI),该模型被看成是第一个真正意义上的主题模型。而Blei等人提出的LDA则在pLSI的基础上进行了扩展,得到一个更为完全的概率生成模型。

主题模型中的一个重要假设是词袋(bag of words)假设,文本被看作是无序的词集合,忽略语法甚至是词的顺序,其维度可能是数万。通过模型的训练,最终聚类得到K个主题<sup>[16]</sup>。

LDA本质上是一种无监督的机器学习模型,将高维词空间表示为低维主题空间,忽略了跟文本相关的标签信息<sup>[17]</sup>。由于LDA模型仅是一个数据降维和聚类算法,所以经常产生不可解释的主题。如果强行将这些主题与文本的标签进行匹配会造成类别判定的精度下降<sup>[18]</sup>。因此,为了解决分类等有监督机器学习问题,人们开始研究有监督学习或半监督学习的LDA模型<sup>[19]</sup>。

半附加类别标签的主题模型(Partially Labeled Dirichlet Allocation, PLDA)就是其中的一个。PLDA与LDA的区别在于,在训练模型时只使用那些与文本可观察的标签集相对应的主题<sup>[20]</sup>。除此之外,还假设有一个共享的、全局的潜在(Latent)主题,它与任何现有的标签没有关联,是作为针对所有文本的背景标签存在的<sup>[21]</sup>。

我们把能力需求特征区分为知识、技能(比如学习能力、解决问题的能力、表达能力等)和素质(比如责任心、积极等),文献回顾中提到的组织技能、非认知技能都属于这里的技能。

针对某些岗位类别,对技能和素质的要求是大同小异的。本文基于PLDA模型对招聘信息进行文本分析的原因在于,希望用这样一个共享的、全局



的潜在(Latent)主题来描述共同的技能和素质。

我们的做法是使用带有标签的预处理后的文本训练PLDA模型,建立起主题与标签之间的一一映射的关系,获得主题-词矩阵和文本-主题矩阵。由于每个主题的种子词反映了能力需求,所以我们可以得到特定类别的岗位的能力需求特征。

### (3) 特征重排序

在上面的主题-词矩阵中,可以发现有的词在各个主题中出现的概率都很高。TF-IDF认为,词的重要性随着它在文本中出现的次数成正比增加,但同时会随着它在语料库中出现的频率成反比下降。所以,需要使用TF-IDF对主题-词矩阵进行重排序。计算公式如为<sup>[22~24]</sup>:

$$\begin{aligned} tfidf_{ij} &= tf_{ij} \times idf_i \\ tf_{ij} &= \frac{\phi_{ij}}{\sum_{k=1}^K \phi_{i,k}} \\ idf_i &= \log \left( \frac{K}{1 + \{j : w_i \in z_j\}} \right) \end{aligned} \quad (1)$$

其中,  $\phi_{ij}$  为主题-词矩阵中词  $w_i$  在主题  $z_j$  中出现的概率。K为主题数,  $\{j : w_i \in z_j\}$  表示包含  $w_i$  的主题数。

经过重排序后,我们选择每个主题的前20个词作为种子词,它们就是特定类别岗位的能力需求特征。

### (二) 设计培养目标和培养规格

人才培养方案是根据专业培养目标和培养规格所制定的实施人才培养活动的具体方案。传统上要确定培养目标和培养规格,或者是广泛开展企业调研,对接职业标准,或者是深入开展校企合作,召开实践专家研讨会。而本文提出的需求驱动的职业能力培养则是在获取某一岗位类别的能力需求特征后,遵循岗位类别——能力需求特征——知识结构、能力结构和素质结构的路径来设计培养目标和培养规格,具体结构如图2所示。

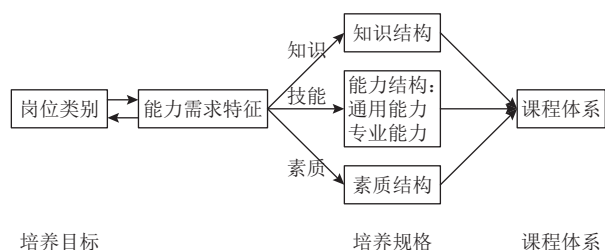


图2 需求驱动的职业能力培养

在明确培养目标时,关键是找准专业人才培养所针对的岗位类别。这是一个反复斟酌的过程,当职业技术学院本身的师资条件满足不了时,需要重新选择岗位类别。

一旦确定岗位类别,就可以按照其能力需求特征设计培养规格——根据知识设计知识结构,根据技能设计能力结构,根据素质设计素质结构。

培养目标与培养规格解决了“培养什么样的人”的问题,而课程体系则是他们的具体化。接下来,职业技术学院可以围绕培养规格构建课程体系,解决“如何培养”的问题。

## 三、案例研究

下面本文以一个实际的案例具体说明如何利用大数据实现需求驱动的职业能力培养。

### (一) 获取岗位的能力需求特征

2016年3月6日我们使用网络爬虫系统从前程无忧抓取了40000条招聘信息,其中岗位类别为软件工程师的有1236条。经过数据预处理,得到了1027条招聘信息的岗位名称语料库和岗位需求语料库。

在文本分析时,首先是打标签。岗位需求语料库的1027个文本中,738个有标签,289个没有标签。接下来,使用带有标签的738个文本训练PLDA模型。标签的种类为30个,但在PLDA模型的参数设置时指定过滤掉只在很少的(<10)文本中出现的标签,最后实际用到的标签是11种,即.net、java、php、前端、android、c、ios、c++、测试、嵌入式、大数据。这些标签对应了PLDA模型中的11个主题。

重排序后得到的11个主题的种子词就是岗位类别为软件工程师的11个方向的能力需求特征。下面我们使用词云对共同的技能和素质、Java方向、前端方向和Android方向进行可视化。

正如前面指出,软件工程师有一些共同的技能和素质需求,包含在一个共享的、全局的潜在(Latent)主题中——比如学习、沟通、经验、独立工作、撰写文档、分析问题、英语、管理、协调等技能和团队精神、责任心、合作、积极、严谨、踏实等素质。另外,还有linux、操作系统、java、数据库、互联网等知识,这些知识往往代表了一个方向或是很多方向共同要求的,如图3所示。

Java方向的基础是java,框架选择j2ee(实际上是架构)或开源的struts、spring、hibernate(持久层也可以是mybatis、ibatis),数据库以oracle和mysql为主,web服务器常用tomcat或jboss,服务器端脚本语言是jsp, servlet用于开发扩展web服务器的应用程序,如图4所示。

前端方向与web开发密切相关,前端的基础是

javascript、html、css, 处理各种浏览器的兼容性是一个重要的方面, 并且要懂布局优化, 流行的框架有jquery、bootstrap和angularjs, ajax使网页实现异步更新, node.js让javascript运行在服务器端, dom是w3c组织推荐的处理xml的标准编程接口, 前端多应用于网站等, 注重用户体验, 了解微信开发更佳, 如图5所示。



图3 共同的技能和素质



图4 Java方向



图5 前端方向

Android方向需要有c和c++基础, 了解网络以及tcp/ip协议和http协议, 擅长ui, 熟悉架构, 强调解决问题和按照编码规范进行开发。除此之外,

android还要求有java基础, 懂socket编程, 使用jni与其他语言通信, 熟悉ndk, 了解各种框架, 会动画, 能够独立工作, 有经验, 善于合作, 如图6所示。

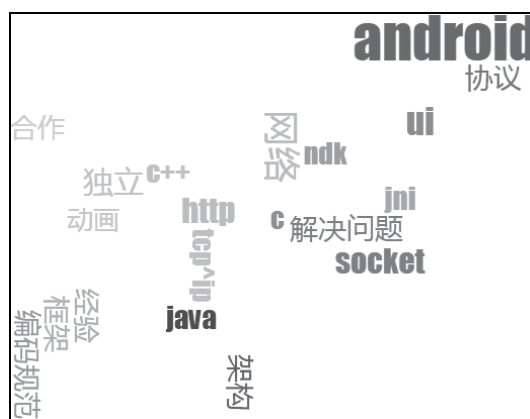


图6 Android方向

## (二) 设计培养目标和培养规格

下面根据以上能力需求特征对四川交通职业技术学院软件技术专业制订的人才培养方案进行修订。

### 1. 问题

软件技术专业创建于2003年, 是四川交通职业技术学院国家示范重点建设的专业之一, 该专业培养学生的目标是成为软件工程师。表1是针对2017级和2018级学生制订的人才培养方案中主要就业的岗位类别、典型工作任务和核心能力。

表1 主要就业的岗位类别、典型工作任务和核心能力

岗位类别	典型工作任务	核心能力
软件开发 (即软件工程师)	1. 模块开发 2. 模块测试 3. 初级技术支持 4. 数据库建立与操作 5. UI界面设计	能够设计与实现简单的B/S型或C/S型系统和移动应用系统

该方案的优点是针对岗位类别和强调核心能力, 缺点是核心能力不够具体。为了将它与软件工程师的能力需求特征进行比较, 我们进一步研究了人才培养方案中对核心课程的描述, 从中发现要求掌握哪些知识点。

软件工程师这个岗位类别下有11个方向, 如果定位不明确, 有可能口径宽但是针对性不强。就人才培养方案目前实施的情况, 我们选择Java、前端、Android方向来比较。

图7是比较的结果。由于人才培养方案是2015年制订的, 沿用至今。所以, 我们可以看到要求掌握的知识点只占了3个方向所需要的知识、技能和素质的一小部分, 如图7中加粗的斜体显示。

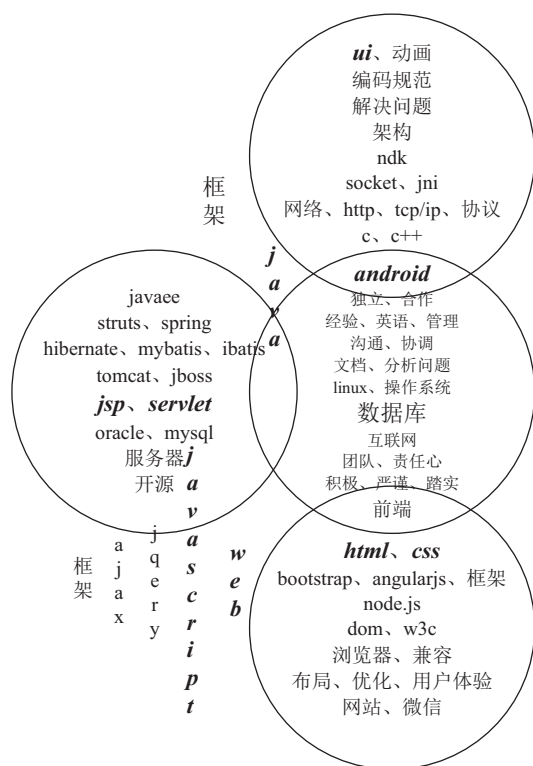


图7 人才培养方案中的问题

## 2. 对策

在目前人才培养方案的实施中,我们也认识到这个问题,增加了开源的框架struts、spring、hibernate的学习,数据库一般为mysql,web服务器常用tomcat。但是,我们的认识仍然与这3个方向的能力需求特征存在一定的偏差。在以下对人才培

养方案的修订中,我们保留Java和Android方向,舍弃前端方向,然后遵循岗位类别——能力需求特征——知识结构、能力结构和素质结构的路径来重新设计培养目标和培养规格。

培养目标: 软件工程师岗位类别下的Java和Android方向。

能力需求特征: 显示在图7中。

知识结构: 根据图7揭示出来的Java和Android方向应该掌握的知识,我们绘制了图8的知识结构(网页设计、系统分析和设计以及数据结构这3个知识是我们增加的)。

能力结构: 通用能力即图3中共同的技能——学习、沟通、经验、独立工作、撰写文档、分析问题、英语、管理、协调等;除了上述通用能力,Android方向的专业能力更强调解决问题和按照编码规范编程。

素质结构: 团队精神、责任心、合作、积极、严谨、踏实等。

在此基础上,可以很容易地把图8中的知识点转换成相应的课程。除了用虚线圈起来的部分(建议设计成包含了多个知识的课程),其他的方框则可以各自对应一门课程。接下来就可以按照内在的逻辑顺序和学生认知规律及职业成长规律,由浅入深、先易后难、先专项后综合、循序渐进地进行排序,构建课程体系。第一学年注重学生专业基础能力培养;第二学年注重学生专业能力训练;第三学年注重学生职业综合能力和创新能力培养。

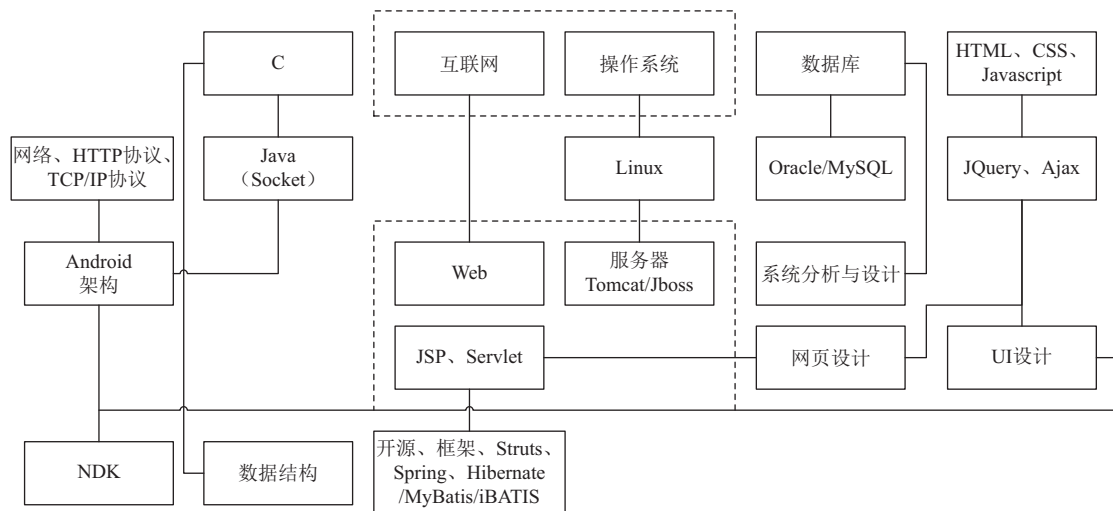


图8 岗位类别为软件工程师（Java和Android方向）的知识结构

职业技术学院以往在制订人才培养方案时需要花费大量的人力、物力,而上面的案例展示了大数

据带来的好处,方案更令人信服、过程更快速、成本更低。

## 四、总结与展望

职业技术学院在修改和更新人才培养方案时,通常小心翼翼、甚至裹足不前,造成的结果就是人才供给与需求的错配。而本文提出的基于大数据的方法,可以帮助职业技术学院实现需求驱动的职业能力培养。该方法有以下两个创新,一是在获取某一岗位类别的能力需求特征时,基于PLDA模拟对招聘信息进行文本分析,同时对建模结果进行重排序,提高了主题种子词的准确性。二是提出了设计培养目标和培养规格的新路径:从岗位类别到能力需求特征,再到知识结构、能力结构和素质结构。本文的案例研究虽然针对的是软件技术专业,但是该方法可以方便地推广到其他专业。

我们也发现招聘信息中的岗位职责能够反映典型的工作任务,所以我们将来的工作是对岗位职责进行文本分析,这样就可以通过深入了解典型的工作任务,完善培养规格中对能力结构的设计。今后我们还要继续研究如何把基于大数据的方法与问卷调查、专家访谈等传统方法相结合,以便更好地为职业技术学院的专业设置和建设提供辅助决策。

## 参考文献

- [1] 金辉.《2016全球人力资本报告》:第四次工业革命催生人力资本新需求[N].经济参考网,2016-06-29.
- [2] 民政部职业技能鉴定指导中心(中民民政职业能力建设中心).现代职业教育体系建设规划(2014-2020年)[N/OL].(2014-06-30).<http://jnjdc.mca.gov.cn/article/zyjd/zcwj/201406/20140600660060.shtml>
- [3] 教育部.普通高等学校高等职业教育(专科)专业目录(2015年)[N/OL].(2015-10-26).[http://www.moe.edu.cn/srcsite/A07/moe\\_953/201511/t20151105\\_217877.html](http://www.moe.edu.cn/srcsite/A07/moe_953/201511/t20151105_217877.html)
- [4] NOMURA S, IMAIZUMI S, AREIAS A C, et al. Toward Labor Market Policy 2.0: The Potential for Using Online Job-Portal Big Data to Inform Labor Market Policies in India[Z]. Policy Research Working Paper, 2017.
- [5] GALLIVAN M J, TRUEX D P, KVASNY L. Changing Patterns in IT Skill Sets 1988-2003: A Content Analysis of Classified Advertising[J]. Acm Sigmis Database, 2004, 35(3): 64-87.
- [6] KUHN P. The Internet as a Labor Market Matchmaker: How Effective are Online Methods of Worker Recruitment and Job Search?[J]. IZA World of Labor, 2014, 18(5): 1-10.
- [7] BEBLAVÝ M, FABO B, LENAERTS K. Skills Requirements for the 30 Most-Frequently Advertised Occupations in the United States: An Analysis Based on Online Vacancy Data[Z]. Ceps Papers, 2016.
- [8] CHAN E S K, SWATMAN P M C. Electronic Commerce Careers: A Preliminary Survey of the Online Marketplace[A]. Proceedings of the 13th Bled Electronic Commerce Conference[C]. Bled, Slovenia: BECC, 2000: 19-21.
- [9] WADE M R, PARENT M. Relationships between Job Skills and Performance: A Study of Webmasters[J]. Journal of Management Information Systems, 2002, 18(3): 71-96.
- [10] CAPILUPPI A, BARAVALLE A. Matching Demand and Offer in On-line Provision: A Longitudinal Study of Monster.com[C]. 12th IEEE International Symposium on Web Systems Evolution, 2010:13-21.
- [11] WOWCZKO I A. Skills and Vacancy Analysis with Data Mining Tech-niques[J]. Informatics, 2015, 2(4): 31-49.
- [12] MAURO A D, GRECO M, GRIMALDI M, et al. Beyond Data Scientists: a Review of Big Data Skills and Job Families[C]. Ifkad: International Forum on Knowledge Asset Dynamics, 2016.
- [13] 程茜, 易烽, 邹奕. 数据挖掘在高校毕业生就业指导上的应用[J]. 科技信息, 2012(10): 233-233.
- [14] 钟晓旭, 胡学钢. 基于数据挖掘的Web招聘信息相关性分析[J]. 安徽建筑工业学院学报(自然科学版), 2010, 18(4): 93-96.
- [15] 彭金莲, 胡祝华, 郑兆华, 陈显毅, 钟杰卓, 李淑. 网络工程专业“3+1”模块化课程体系的研究[J]. 海南大学学报(自然科学版), 2013, 31(1): 74-79.
- [16] 徐戈, 王厚峰. 自然语言处理中主题模型的发展[J]. 计算机学报, 2011, 34(8): 1423-1436.
- [17] 张群, 王红军, 王伦文. 词向量与LDA相融合的短文本分类方法[J]. 现代图书情报技术, 2016, 32(12): 27-35.
- [18] 江雨燕, 李平, 王清. 用于多标签分类的改进Labeled LDA模型[J]. 南京大学学报(自然科学版), 2013, 49(4): 30-37.
- [19] 郑世卓, 崔晓燕. 基于半监督LDA的文本分类应用研究[J]. 软件, 2014(1): 46-48.
- [20] RAMAGE D, HALL D, NALLAPATI R, et al. Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-labeled Corpora[C]. Singapore: Conference on Empirical Methods in Natural Language Processing, 2009: 248-256.
- [21] RAMAGE D, MANNING C D, DUMAIS S. Partially Labeled Topic Models for Interpretable Text Mining[C]. 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2011: 457-465.
- [22] 寇宛秋, 李芳. 基于种子词汇的话题标签抽取研究[J]. 中文信息学报, 2013, 27(5): 114-121.
- [23] 闫泽华. 基于LDA的新闻线索抽取研究[D]. 上海: 上海交通大学, 2012.
- [24] SONG Y, PAN S, LIU S, et al. Topic and Keyword Re-ranking for LDA-based Topic Modeling [C]. Hong Kong: ACM Conference on Information and Knowledge Management, 2009: 1757-1760.

编辑 何婧