

Analysis and Mining of Network Recruitment Information

Dongdong Wang, Xiaoyan Zhou, Jingjing Xiang, Xiaomin Guo, Qun Wang

Department of Mathematics and Computer Science, Guizhou Normal College, Guiyang Guizhou
Email: 674714665@qq.com

Received: Dec. 8th, 2016; accepted: Dec. 22nd, 2016; published: Dec. 29th, 2016

Copyright © 2016 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The employment and network recruitment information is based on the study of the era of big data problems through the R language analysis of the Network Recruitment Information and text mining, mainly from the city, industry, and several aspects of this position. The analysis is related to the current situation of talent demand, forecasts the number of posts and provides job seekers with grey number of the prediction model, accomplishing the enterprise and talent performance in network platform interaction. This can achieve better function. Finally, the thesis will put forward proposals to the relevant corporate recruitment and employment planning guidance.

Keywords

Recruitment Information, Text Analysis, Word Cloud, Grey Prediction

网络招聘信息的分析与挖掘

王冬冬, 周小燕, 向晶晶, 郭小敏, 王 群

贵州师范学院, 数学与计算机科学系, 贵州 贵阳
Email: 674714665@qq.com

收稿日期: 2016年12月8日; 录用日期: 2016年12月22日; 发布日期: 2016年12月29日

摘 要

本文以研究大数据时代下人才就业与网络招聘信息之间的问题为基础, 主要从城市、行业、职位这几个

方面通过R语言对网络招聘信息进行文本分析与挖掘，分析目前的人才需求情况，并用灰色预测模型预测出提供岗位数与求职人员数，实现企业与人才间在网络平台上的互动，达到更好的匹配功能，最后给相关企业的人才招聘以及大学生的就业规划指导提出自己的建议。

关键词

招聘信息，文本分析，词云，灰色预测

1. 引言

信息时代，我国互联网得到迅速发展，网络人才市场势头迅猛，严重冲击了传统的人才市场，网络人才市场占有率不断扩大，人才招聘渠道呈现多元化的发展，网络招聘作为网络技术人力资源管理中应用最快的领域，已经被人们所接受，据统计全球每天约有 2000 万就业信息发布，约有 3000 万人在互联网上发出求职简历。网络招聘的发展促进了网络招聘市场的繁荣，它与传统招聘渠道相比有着更为明显的优势。网络招聘信息一方面能够直接反映用人单位对人才基本条件、能力和素质的要求，为应聘者提供求职参考；另一方面也能够反映出社会和各行业对人才的需求现状，或未来一段时间的人才需求趋向；同时也能够为高等院校及时了解社会对人才的需求变化情况，分析预测未来的人才市场的热点，有针对性的调整人才培养方案和设置安排相关课程提供重要的参考信息，促进高校培养出更多适用的优秀人才以满足社会的需求。因此，对网络招聘信息进行分析研究，了解社会和相关行业的需求特点与趋势，使企业在人力资源招聘中节省成本、招聘到企业实际需要的人员，招聘网站能更好的实现企业与人才间的互动，达到更好的匹配功能，为广大求职者提供正确的就业指导都有着非常重要的意义。

2. 分析过程

2.1. 数据来源

本文数据来自于国内招聘网站(赶集网, 58 同城, 前程无忧网) 2016 年 1 月到 3 月的招聘信息以及中国人力资源市场网 2012 年第一季度至 2016 年第一季度的招聘信息。其中城市、行业和职位数据属于非结构性数据。

2.2. 数据处理

对数据进行预处理[1]，先把异常值和缺失值筛选、整理，再把处理后的数据进行词频统计，这里保留词频大于 100 的数据，因为 100 以下的所占比重较小，可以忽略，简化数据，降低分析难度。

2.3. 数据分析

对收集到的招聘信息数据中的城市，行业，职位这三列数据进行词频统计，进行频数统计分析，通过频数分布表、条形图、地域分布图，以及集中趋势和离散趋势的各种统计量来描述数据的分布特征，可以直观的得到，热门城市间差距的大小、热门行业以及热门职业间竞争的相对大小，同时通过 R 语言分别画出它们的词云[2]，在词云图中，通过字体的大小判断热门城市、行业及职位。

表 1 是对前十名热门城市的词频统计，可以直观的看出城市间热门差异的大小。

图 1 是热门城市的词云统计图，明显的看出北京、上海、深圳、广州、杭州这几个城市比较突出，由此我们可以得到这几个城市在中国是热门的城市，而且和上面的词频统计结果也是吻合的。

从图 2 可以很明显的看到全国的就业热门城市基本上集中在东部地区，而且主要集中在沿海城市。

Table 1. Statistics of frequency city
表 1. 十大热门城市词频统计表

城市	北京	上海	深圳	广州	杭州	成都	武汉	南京	厦门	西安
频数	197,258	88,408	64,516	46,370	41,654	15,521	9960	8559	5901	4950



Figure 1. Popular city statistics
图 1. 热门城市统计图

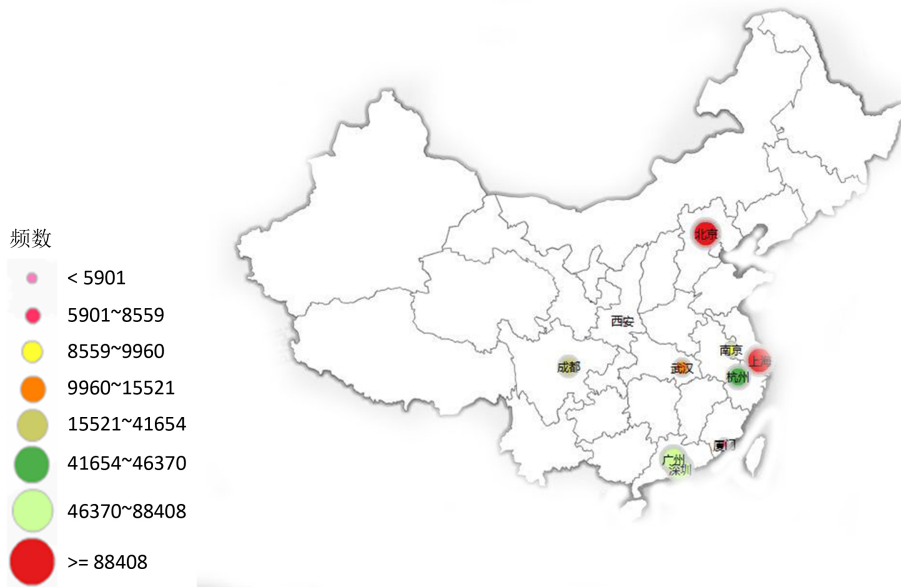


Figure 2. Regional distribution of hot cities
图 2. 热门城市区域分布图

表 2 是热门行业的词频统计，由上可知在中国的就业市场中移动互联网·电子商务、金融、移动互联网·O2O 等是非常热门的行业，同时可以得到行业间竞争的激励程度，从表 2 得到移动互联网·电子商务的频数为 42,227，金融的频数为 39,410 而移动互联网·O2O 的频数为 29,997 由此可见移动互联网·电子商务和金融行业的人才需求量特别大。

图 3 是对各城市前三的行业的一个对比，广州、杭州、等的第三热门行业未在图中显示(数据太小，图上显示不明显)，从图中可以很明显的得到，北京不仅是最热门的城市，而且他的各个行业的热度相比于其他城市同类行业也处于领先地位。

Table 2. Statistics of hot industries' word frequency
表 2. 热门行业词频统计表

行业	频数
移动互联网·电子商务	42,227
金融	39,410
移动互联网·O2O	29,997
电子商务	29,752
移动互联网·金融	25,434
移动互联网·数据服务	17,444
移动互联网·企业服务	15,489
移动互联网·教育	12,740
移动互联网·游戏	12,366

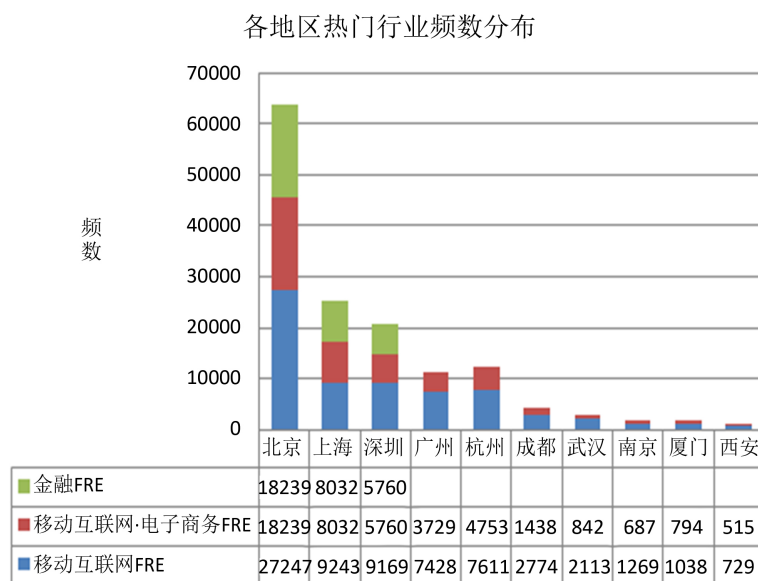


Figure 3. Hot industry distribution
图 3. 热门行业分布图

图 4 是反映职业的词频统计[3]，也可同上面做类似的分析，得到产品经理和 Java 这两个职业的热门程度不相上下。

由图 5 可以直观的得到：在中国的就业市场中产品经理、Java、web 前端、等职业的需求量比较大。

从中国人力资源市场网收集了 2012 年第一季度到 2016 年第一季度全国各单位所提供岗位数和全国求职人数，对全国的供求结构进行下面的分析，并对其做预测[4]。

通过 excel 对 2015 年第三季度用人单位对学历的需求数作雷达图，形象的描绘出不同学历的需求情况。从图 6 中可以看出，需求数最高的是大专其次才是本科，且对高学历人才的需求数相对较少。由此可以得出结论：我国现阶段的大部分职位高学历人才并不多，反而是以专科等技术性人才为主，这些高校培养的是技能型人才，毕业后，大部分都从事生产性岗位，相比本科生来说，他们的技能可以更直接的运用到生产工作中，但这也很可能导致部分企业形成技术落后、市场竞争力不强、经济效益差、人才引进难等恶性循环的局面。虽然企业对求职人员学历的要求并不高，但并不代表企业不需要高学历人才，

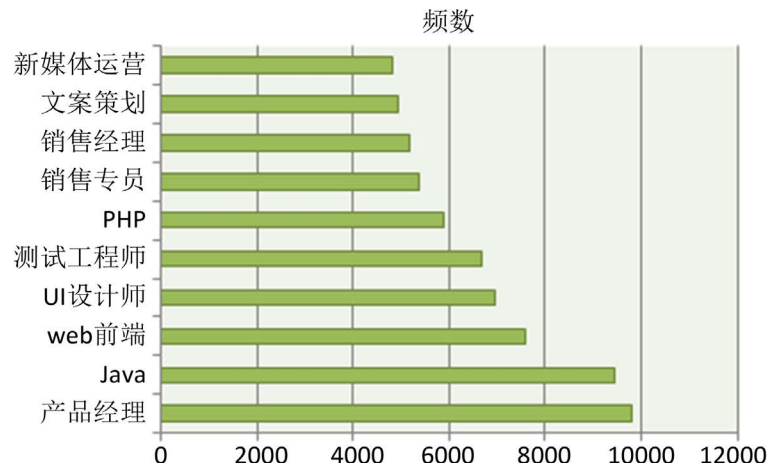


Figure 4. The frequency of popular career statistics
图 4. 热门职业的频数统计图

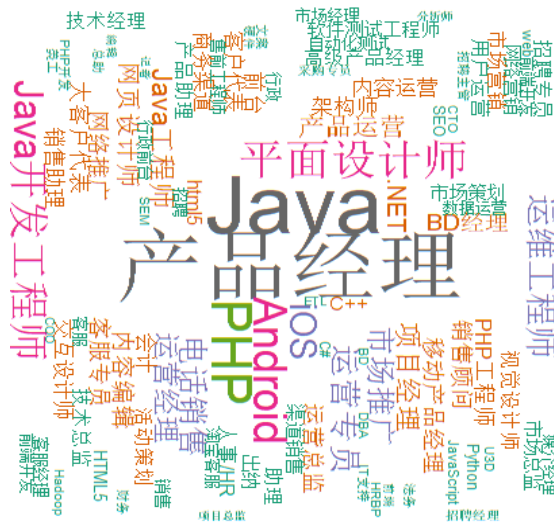


Figure 5. Popular career statistics
图 5. 热门职业统计图

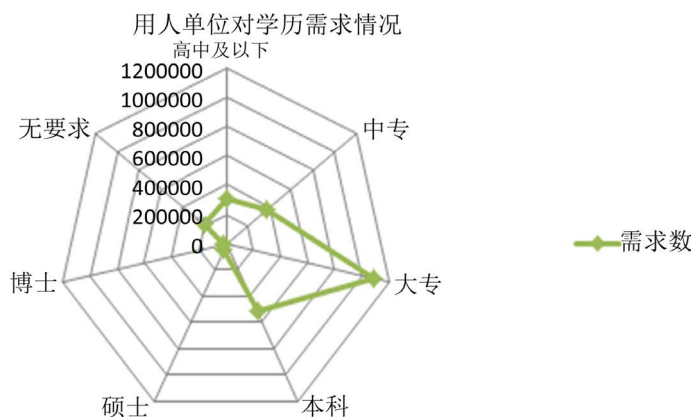


Figure 6. Radar chart
图 6. 雷达图

由于我国中小型企业占全国企业总数的 99% 以上，其中一大部分企业因为人才成本过高且本身从事制造业，他们更愿意用学历较低的廉价劳动力，但实际上，企业还是很缺乏高学历人才的。

图 7 是反映的不同地区的市场供求总体状况气泡图，圈大小表示供求需要的数量，从图中可以看出，华东地区的人才供求量最大，其次是华南、华北等地区。而西北、东北地区的供求量相对偏低。说明岗位的需求大部分集中在了华东、华南等地区，这些地方的经济发展水平较高，相应的人才需求也就越大。

为了预测全国供求未来发展趋势的状况，我们可以通过灰色预测模型来预测求职人数与提供岗位数。灰色预测是一种对含有不确定因素的系统进行预测的方法，通过鉴别系统因素之间发展趋势的相异程度，即进行关联分析，并对原始数据进行生成处理来寻找系统变动的规律，生成有较强规律性的数据序列，然后建立相应的微分方程模型，从而预测事物未来发展趋势的状况。用等时距观测到的反应预测对象特征的一系列数量值构造灰色预测模型，且认为原始数据本身就是一种多因素综合作用的结果，与其进行因素的多层剖析，不如就以原始数据为依据进行预测。故用单因素模型 GM(1,1) 作预测[5]。

GM(1,1)模型的理论[6]

令 $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), x^{(0)}(3) \dots, x^{(0)}(n))$ 为原始建模序列

作一次累加生成 k

$x(k) = \sum x^{(m)}$ 消除数据的随机性和波动性

$m = 1$

有 $x^{(1)} = (x^{(1)}(1), x^{(1)}(2), x^{(1)}(3) \dots, x^{(1)}(n))$
 $= (x^{(1)}(1), x^{(1)}(1) + x^{(1)}(2), \dots, x^{(1)}(n-1) + x^{(1)}(n))$

x 可建立白化方程： $dx/dt + ax = u$ 即 GM(1,1)。

该方程的解为： $x(k+1) = (x(1) - u/a) \exp(-ak) + u/a$ 其中： a 称为发展灰数； μ 称为内生控制灰数。用 R 语言进行编程得到结果如图 8 所示。

求职人数拟合函数：

$$x(t+1) = 31640.87 \exp(-0.02937302) - 31640.87$$

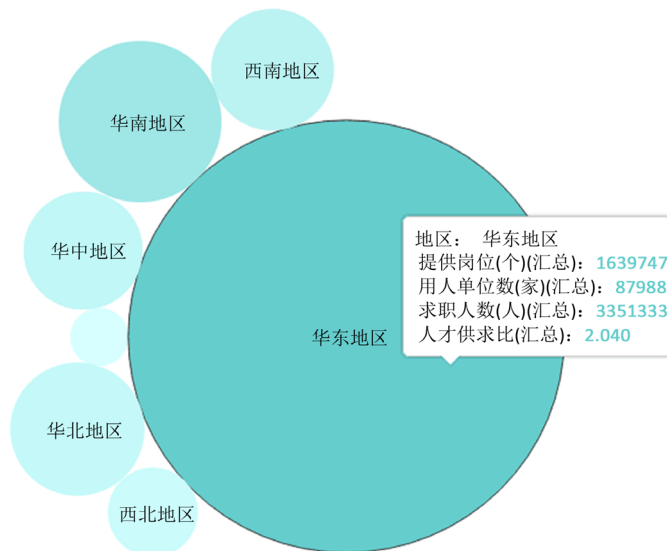


Figure 7. Bubble chart
图 7. 气泡图

最终下一期求职人数大约为 560 万人。由于 $C(\text{后验差比值}) < 0.35$ ，说明 GM(1,1)模型预测[7]精度等级为好。相对精度为 91.91678%，说明这个模型的精度非常高，具有一定可靠性(图 9)。

拟合函数：

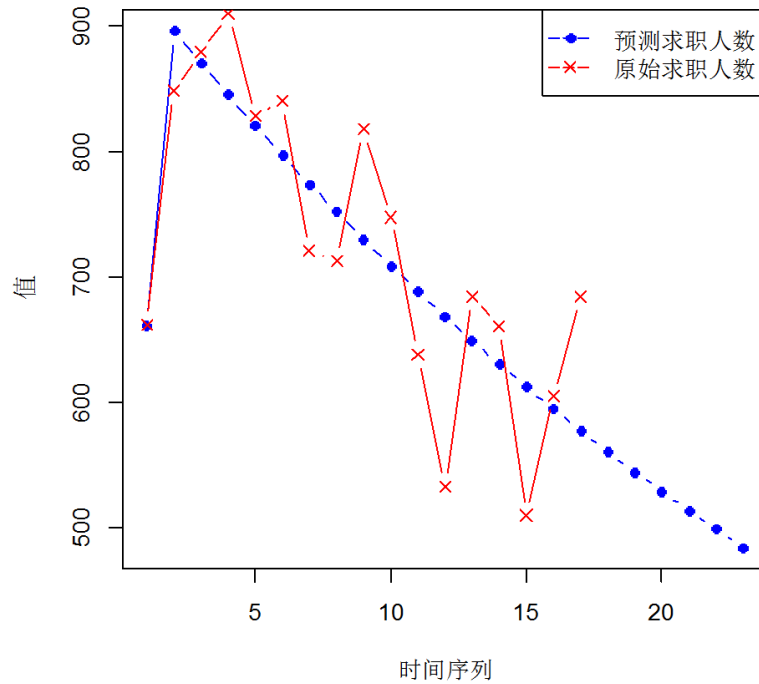


Figure 8. Job applying forecast chart
图 8. 求职预测图

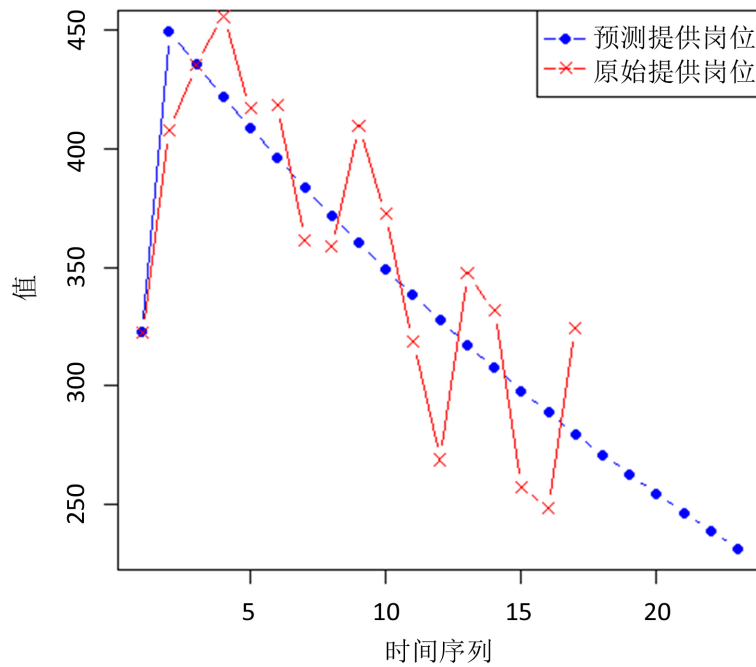


Figure 9. Position forecast
图 9. 岗位预测

$$y(t+1) = 14778.48 \exp(-0.03158537) - 14778.48$$

最终预测下一期的提供岗位数大约为 271 万, 由于 C 值 < 0.35 , 说明 GM(1,1)模型预测精度等级为好。相对精度为 91.07493%, 同样说明这个模型的精度非常高, 具有一定可靠性。

从预测情况看出, 从 2016 第一季度开始未来六个季度劳动力市场需求大于供给。从供求对比看, 用人单位提供岗位数约为 271 万。进入市场的求职者约 560 万人, 市场需求大于供给。

3. 结论

通过以上的分析, 我们得出北京、上海、广州等是中国的热门就业城市, 而移动互联网·电子商务、金融、移动互联网·O2O 等是排名靠前的行业, 同时产品经理、Java、web 前端等的人才需求量很大。我国现阶段大部分职位以技术性人才为主。由本文预测结果来看, 劳动力市场需求大于岗位供给, 就业形势仍然严峻。而且现在的热门行业中需要的是高端技术人才, 互联网是最近几年逐步发展起来得, 很多学生所学专业与该行业不符, 这也是就业难的一个原因。造成“企业招工难, 大学生就业难”的尴尬局面。基于本文研究结果提出下列建议:

1) 高校应该适当增加互联网领域相关的专业, 或对这些专业加大资源投入和调整课程设置加强对学生专业技能的训练。

2) 从人才类型来看, 需求量大的是技术型人才, 对于应用型学校可请企业技术人员到学校做讲座, 加强校企合作, 提高学生就业率和综合素质能力, 培育出适合未来就业形势的高技能创新型人才。

3) 大学生在做职业规划时应多从这些热门行业上做准备。

4) 大学生应该准确的认知自我, 不能盲目随主流, 从基层做起, 在择业时多考虑中部与西部的二线城市, 缓解就业压力。

基金项目

贵州师范学院校级学生科研项目(项目编号: 2016DXS094); 贵州省 2014 年省级本科教学工程项目“计算机科学与技术”专业综合改革(项目编号: 黔教高发[2014] 378 号); 卓越工程师教育培养计划项目(黔教高发[2013] 446 号); 2015 年省级本科教学工程建设项目(黔教高发[2015] 337 号)。

参考文献 (References)

- [1] 李妍. 微博数据预处理及话题检测方法研究[D]: [硕士学位论文]. 石家庄: 河北师范大学, 2014.
- [2] 王成城, 孙丽丽, 朱立萍, 巩妙宇. NGO 组织人力资源管理的困境与对策——基于网络招聘信息的文本分析[J]. 中国人力资源开发, 2015(17): 20-25.
- [3] 钟晓旭. 基于 Web 招聘信息的文本挖掘系统研究[D]: [硕士学位论文]. 合肥: 合肥工业大学, 2010.
- [4] 王立敏. 人才市场需求预测模型的研究与应用[D]: [硕士学位论文]. 石家庄: 河北师范大学, 2013.
- [5] 居柏成. 基于灰色理论的大学生就业市场信息数据挖掘研究[D]: [硕士学位论文]. 长春: 东北师范大学, 2006.
- [6] 百度百科. 灰色预测[Z/OL]. <http://baike.baidu.com/link?url=tCdFYbL3eShziNnz0Uq1PSII2FqIDVIXr0S1ri3LV7TG3Ineza9tcn0yL3MEUBblJjN4Rp2RIkp8G9Z9xx01q#5>
- [7] 李梦婉, 沙秀艳. 基于 GM(1,1)灰色预测模型的改进与应用[J]. 计算机工程与应用, 2016, 52(4): 24-30.

期刊投稿者将享受如下服务：

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：sa@hanspub.org