

基于 ECHARTS 的现代远程教育招生数据 交互式可视化研究及实现

纪娟^{1,2} 彭扬¹

(1. 四川广播电视大学, 成都 610073;

2. 国家开放大学教育信息管理与信息系统研究中心, 成都 610073)

摘要: 招生数据是所有教育数据的基础来源。针对通用的数据可视化图表难以体现国家开放大学招生数据的分布特征, 结合办学机构的区域性和招生时间的时效性, 从时间维度和度量指标设计招生数据交互式可视化流程, 采用 AJAX 技术异步加载全国和四川分部近 4 期招生数据, ECHARTS 技术实现了招生数据交互式可视化, 最终以可视化的形式直观地展现给决策层, 为其提供决策依据。

关键词: 可视化; 招生数据; 交互式; ECHARTS

中图分类号: TP319

文献标识码: A

文章编号: 1671-6558(2019)03-28-04

DOI: 10.3969/j.issn.1671-6558.2019.03.006

Research and Implementation of Interactive Visualization of Modern Distance Education Enrollment Data Based on ECHARTS

JI Juan^{1,2} PENG Yang¹

(1. Sichuan Radio and TV University, Chengdu 610073, China;

2. Educational Information Management and Information System Research Center
in the Open University of China, Chengdu 610073, China)

Abstract: Enrollment data is the basic source of all educational data. Aiming at the difficulty of visualizing the distribution characteristics of enrollment data in National Open Universities by using general data visualization charts. Combining with the regionality of school-running institutions and the timeliness of admission time, this paper designs an interactive visualization process of enrollment data from the time dimension and measurement indicators, and asynchronously loads the latest four phases data of national and Sichuan branch with AJAX technology. ECHARTS technology realizes interactive visualization of enrollment data, and finally, the visual form is presented intuitively to the decision-making level to provide the basis for decision-making.

Key words: visualization; enrollment data; interactive; ECHARTS

收稿日期: 2019-06-04

基金项目: 四川广播电视大学重点科研项目(KTZXJY2018002Z); 四川广播电视大学教改项目(XMGCJS2016007Y)。

作者简介: 纪娟(1983-), 女, 安徽安庆人, 工学硕士, 讲师, 主要从事信息系统开发工作。

0 引言

社会经济和信息技术的快速发展推动了大数据时代的到来,挖掘大数据间所蕴涵的知识在社会经济进步和科学技术发展领域有着重大战略意义^[1]。自2011年世界著名咨询管理麦肯锡公司发布关于“大数据”的报告后,“大数据”获得各行各业高度重视和关注。当前,我们的世界已经进入大数据时代^[2]。“大数据”只是一个概念,而利用数据挖掘和数据分析相关技术去探索“大数据”背后潜藏的价值,最终把数据转化为可视化的信息直观地展现给决策者才是真正的关注点。“大数据”的内涵研究是数据挖掘、数据分析和数据可视化。数据挖掘和数据分析是过程,数据可视化是结果,挖掘和分析的最终目的是把可视化的信息展现给决策者。因此,数据可视化是大数据处理中必要的不可缺少的部分。

教育数据内涵广泛^[3],招生数据、学籍数据、学习数据、成绩数据和毕业数据基本涵盖了教育数据的范围,在这5大类型数据中招生数据是所有数据的基础来源,也是高校领导者们最重视的部分,特别是对于依托信息化系统办学的现代远程教育学校来说,生源是生命力。基于此,对招生数据进行分析并可视化是非常必要的,也具有实用价值。

1 数据可视化技术相关研究

1.1 数据可视化软件和技术

大数据时代实现数据可视化的软件很多,选择出适应各自需求的软件必能事半功倍。目前有3种实现方法。

(1) 通过开源和商业软件实现数据可视化。开源软件可免费使用,如 Excel。在 Excel Power View 未出现以前,需要有强大的 VBA 宏编程能力才能在 Excel 中完成数据展示,Excel 2013 及之后版本的出现,可直接连接数据库获取数据源。商业软件收费使用,如 Tableau、FineBI 等。这两款软件功能强大,并支持数据引擎、即时分析、多维度分析和 Dashboard 等功能。

(2) 通过前端框架技术实现数据可视化。这些框架基于前端浏览器 Canvas 和 SVG 渲染技术实现数据可视化,如 HighCharts.js、Echarts.js 和 D3.js。其中,HighCharts.js 和 Echarts.js 是同一种类型的框架,前者是国外知名的商业框架,有详细的英文功能和 API 文档;后者是国内百度团队出品的开源框架,实例和 API 文档较多;D3.js 灵活性高、数据计算和处理、数据映射到图形能力强、可定制并可按创造力绘制图形。

(3) 通过数据科学方面的编程语言原生态地实现数据可视化。计算机软件编程语言较多,如 JAVA、C++、PHP、Python 和 R,而适合于数据可视化和数据分析需求的编程语言只有 Python 和 R,两者都有强大的数学计算算法。

以上3种方法均可实现数据可视化,第1种方法中使用数据可视化软件实现的可视化是静态的图片,无法直接实现人和数据的互动,第3种方法中使用原生态的编程语言实现数据可视化,专业性太强,更适合于有数据挖掘和数据分析需求的项目。因此,为使数据可视化具有交互性,综合因素考虑本文选择第2种前端框架技术中的 ECharts.js 技术实现数据可视化。

1.2 ECharts.js 简介

ECharts 是由百度公司推出的一款非常优秀的可视化图表控件,底层依赖 Canvas 类库 ZRender^[4],并且百度 ECharts 全部免费^[5]。ECharts 可以流畅地运行在 PC 和移动设备上,兼容当前绝大部分浏览器,能够提供直观、生动、可交互、可高度个性化定制的数据可视化图表。在国内应用领域广泛,提供了丰富的可视化图形,并支持多种图形混合使用,交互性强,具有方便使用、定制灵活等优点,通过增量渲染技术,配合各种细致的优化,ECharts 支持千万级的数据量的可视化,并且支持千万级数据量流畅稳定的交互式可视化。

2 基于交互式的学生数据可视化设计

交互是指用户对可视化图表发出某种指令后,图表给出相应的响应,如缩放、拖动和单击等。交互式可视化让数据更加鲜明和生动。本文通过增加时间、度量指标和指令操作使招生数据的可视化具有交互性。招生数据交互可视化设计流程图,如图1所示。

2.1 时间维度

不同于普通高校的招生时间,远程教育学校一年可以在春季和秋季进行招生,招生数据具有年度和季节属性。因此,在数据可视化中添加时间维度是可行的。时间维度信息可视化,通常使用时间轴来实现,时间轴是查看指标值变化趋势的最佳表现形式。本文选取近2年的招生数据,在折线图可视化中添加时间轴,时间轴上有4个节点,分别是2017年春季、2017年秋季、2018年春季和2018年秋季。用户在电脑设备上查看数据时通过点击时间轴上的时间节点实现交互,在移动设备上通过滑动指令互动。

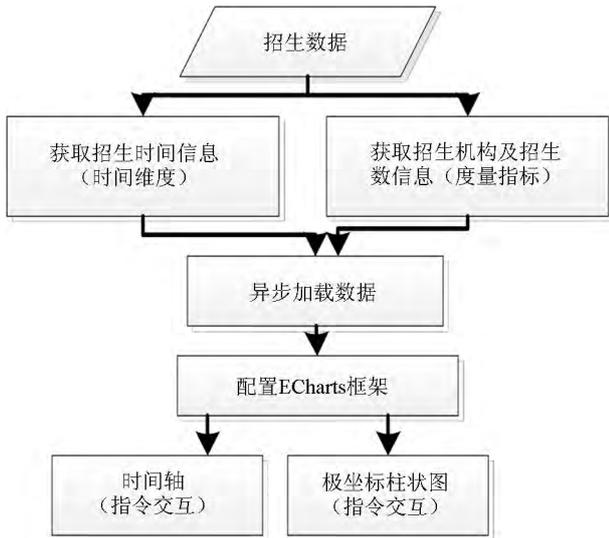


图 1 招生数据交互式可视化设计流程图

2.2 度量指标

国家开放大学是层级办学模式,在管理上逐级接受上一层的管理,主要由分部(省级)、学院(市级)、学习中心(县级)3个部分机构组成,学习中心负责实质性和具体性事物,如招生、组织考试等。可见,招生数据具有空间上的度量特性,表示规模和数量的可视化,通常使用柱状图来实现。本文选取全国各分部招生数据和四川分部下的各学院招生数据,在时间线上及极坐标系内以柱状图形式表示招生数据。从各分部的柱状图点击进入各学院招生数据可视化,以柱状图可视化;从各学院的柱状图点击进入各学习中心招生数据可视化,以柱状图可视化。

3 招生数据交互式可视化实现

3.1 准备数据

数据是可视化过程中必不可少的元素,任何类型可视化过程中的第1步是数据的收集和处理。主要包括:收集数据、设计数据和转换数据。数据收集和處理流程图如图2所示。

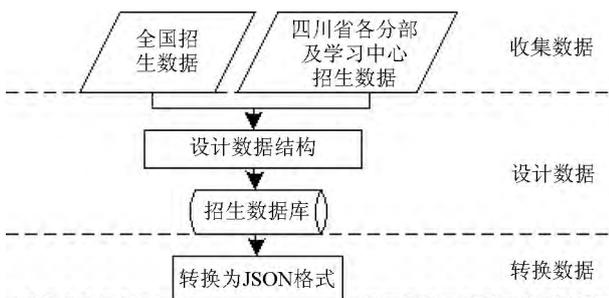


图 2 招生数据交互式可视化数据收集和处理流程图

3.1.1 收集数据

本文需要收集全国、四川各学院及学习中心3个部分的招生数据。

全国招生数据由于基础数据库部署在总部,收集这一部份的数据相对比较繁杂,需要将每期总部下发的招生数据 Excel 文件按照年度、学期、学校和招生数格式汇总成 1 个表。以 2018 年春季为例,部分分部招生数据如表 1 所示。

表 1 2018 春部分分部招生数据表

年度	学期	学校代码	学校名称	招生数
2018	春季	110	北京广播电视大学	8 572
2018	春季	120	天津广播电视大学	13 244
2018	春季	130	河北广播电视大学	23 364
2018	春季	140	山西广播电视大学	10 807
2018	春季	150	内蒙古广播电视大学	14 924
2018	春季	210	辽宁广播电视大学	16 980
2018	春季	211	沈阳广播电视大学	6 472

四川各学院及学习中心招生数据由于基础数据库部署在各分部,数据库有专业技术人员维护和管理,对数据库字典非常熟悉,可快速按需求收集数据。需收集的数据属性有年度、学期、学校代码、学校名称和学号。学校代码中前 5 位数表示学院代码,全部 7 位数表示学习中心代码。因此,学院和学习中心招生数可以共用 1 张带有学生信息的数据作为基础表。以 2018 年春季为例,部分学院和学习中心数据基础表如表 2 所示。

表 2 2018 春四川某学习中心基础数据表

年度	学期	学校代码	学校名称	学号
2018	春季	5100300	四川电大直属学院城铁分院	1851001200001
2018	春季	5100300	四川电大直属学院城铁分院	1851001200002
2018	春季	5100300	四川电大直属学院城铁分院	1851001200003
2018	春季	5100300	四川电大直属学院城铁分院	1851001200004
2018	春季	5100300	四川电大直属学院城铁分院	1851001200005
2018	春季	5100300	四川电大直属学院城铁分院	1851001200006
2018	春季	5100300	四川电大直属学院城铁分院	1851001200007

3.1.2 设计数据

设计数据是数据准备环节中的重要步骤,包括 2 个部分的内容:一是对表 1、表 2 基础数据数据结构的设计,二是招生机构与行政区域信息数据的关联设计。其中关键点在于招生基础库中分部、学院和学习中心分别与行政区域中省、市和县的对应关系的设计。招生机构与行政区域数据的关联设计的数据结构如表 3 所示。

表 3 招生机构与行政区域信息数据的关联设计数据结构

代码	名称	数据类型	注释
Id	行政区域 Id	Int	省、市或县的 id
Name	行政区域名称	varchar(100)	省、市或县的名称
xxdm	学校代码	varchar(50)	分部、学院或学习中心代码
xxmc	学校名称	varchar(200)	分部、学院或学习中心名称
fid	父级	Int	表示层级
bz	备注	varchar(50)	

$x_2 \cdot x_{31} \cdot x_2 \cdot x_{32}$];

[b ,bint r ,rint ,stats] = regress(log(y) ,a);

得到模型的检验统计量 stats 及回归系数 b 如

下:

stats =0. 9871 496. 2196 0. 0000 0. 0011

b =9. 2663 0. 0297 0. 2042 0. 0876 -0. 0020

0. 2132 0. 0561

此时 stats 的第 3 个值接近 0 ,表明回归模型成立 第 1 个值 0. 9871 接近 1 ,表明模型的拟合优度好 这说明软件行业员工的薪资与年限、岗位、学历显著相关 ,且工作岗位与学历有显著的交互影响 ,其对应关系可用如下模型来刻画:

$$\log(y) = 9.2663 + 0.0297x_1 + 0.2042x_2 + 0.0876x_{31} - 0.0020x_{32} + 0.2132x_2 \cdot x_{31} + 0.0561x_2 \cdot x_{32} \quad (7)$$

据此模型可以分析和预测软件行业员工的薪资状况 ,具体来说其作用表现如下:

(1) 帮助毕业生预估入职薪资。例如某本科毕业生应聘技术岗 ,将年限 $x_1 = 0$,岗位 $x_2 = 0$,学历对应逻辑变量 $x_{31} = 1$, $x_{32} = 1$,代入模型 (7) ,可得其薪资期望值为 11521 元 ,即月薪期望值约为 1.15 万。

(2) 帮助公司制定薪资方案。假设学历和岗位都相同时 ,考察不同工作年限薪资的差异 ,利用模型 (7) 可得逐年工资的平均涨幅约为 5% ,例如对新入

职员工而言 ,每年月薪增长约为 600 650 720 等等;假设学历和年限都相同 ,以工作 5 年为例 ,管理岗的月薪比技术岗高约 50%;假设年限和岗位都相同 ,考察不同学历 ,本科生月薪比大专生高约 9% ,研究生比本科生高约 0.3% ,表明不同学历间的工资差异不大。

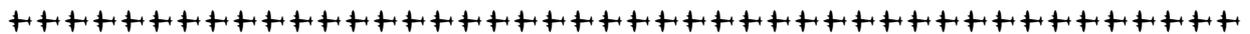
4 结论

本文基于对某软件公司 57 名员工薪资的调查数据 ,建立了多元线性回归模型研究了员工薪资与工作年限、工作岗位、学历三者之间的关系 ,模型不仅通过了检验而且回归系数的拟合优度较高 ,能够帮助求职者预估薪资及涨幅情况 ,也可以为软件公司制定合理的薪资方案提供理论依据。

参考文献

- [1]天极数据调查中心. IT 行情调查研究报告 [M]. 北京: 清华大学出版社 2015.
- [2]何晓群. 应用回归分析 [M]. 北京: 中国人民大学出版社 2015.
- [3]王倩 ,加春燕. 数学建模方法与应用 [M]. 北京: 北京师范大学出版社 2016.
- [4]王松桂. 线性回归与方差分析 [M]. 北京: 高等教育出版社 ,1999.
- [5]姜启源 ,谢金星. 实用数学建模 [M]. 北京: 高等教育出版社 2014.

(责任编辑: 张艳霞)



(上接第 31 页)

4 结论

本文分析了基于 ECHARTS 数据可视化的关键技术 ,在时间维度和度量指标上设计了基于 ECHARTS 的招生数据交互式可视化的流程 ,介绍了时间轴、柱状图等关键图形的绘制方法 ,在此基础上异步加载全国和四川分部近 4 期招生数据 ,并实现了基于 ECHARTS 的数据交互式可视化 ,最终以可视化的形式直观展现给决策层为其提供决策依据。但是 ,本文只收集到四川所属学院及四川所属学习中心的招生数据 ,数据在空间维度上存在着不足 ,数据可视化在交互性上体验友好度降低。在以后的工作中 ,可进一步获取其他分部所属学院及学习中心招生数据 ,增加数据的空间维度 ,提高交互式数据可视化的体验友好度。此外 ,由于开放教育办学机构的开放性 ,使得开放教育办学机构与行政地图中的行

政区域不是一对一的对应关系 ,如在四川开放教育有四川分部和成都分部 2 家办学机构 ,使得本文数据可视化存在着不全面性 ,在后续的研究中可考虑从数据层面上改进和优化。

参考文献

- [1]刘文远 ,李少雄 ,王晓敏 ,等. 大数据知识发现 [J]. 燕山大学学报 2014 ,38(5) : 377-380.
- [2]MCAFEEA ,BRYNJOLFSSONE. Big data: the management revolution [J]. Harvard Business Review ,2012 ,90(10) : 60-68.
- [3]胡祖辉 ,施佳. 高校学生上网行为分析与数据挖掘研究 [J]. 中国远程教育 2017(2) : 26-32.
- [4]王子毅 ,张春海. 基于 ECharts 的数据可视化分析组件设计实现 [J]. 微型机与应用 2016 ,35(14) : 46-48.
- [5]徐欣威. 基于 ECharts 的科技统计数据可视化设计与实现 [J]. 天津科技 2019 ,46(3) : 66-70.

(责任编辑: 张艳霞)