针对专业分流数据的双层放射环矩阵可视化

李 慧[1] 陈红倩[2] 董 爽[1] 马丽仪[1]
(北京联合大学管理学院 北京 100101)[1] (北京工商大学计算机与信息工程学院 北京 100048)[2]

摘 要 为表达并分析专业分流数据中的多属性流向特征分析,提出了一种双层放射环矩阵的可视化方法。该方法首先基于各属性条件对专业分流数据进行筛选和分类统计,然后将各统计结果指标映射为可视元素并展现到最终的可视化结果中。可视元素映射过程主要包含 3 部分:借鉴气泡图方法,将专业分流中的学生人数映射为气泡图形式;将饼图中的对比关系引入气泡图,将专业分流中各条件下的性别比例映射为放射环内层的饼图形式;借鉴 Sunburst 方法,将不同绩点条件下的学生数统计结果映射为放射环外层的区块。依托案例数据的实验结果与学生管理人员的评测结果表明,该可视化结果能对学生人数、性别、绩点等属性的流向特征进行直观表达,并能够对专业分流中的学生进行快速分类与细化,对专业建设与培养计划的制定起到良好的指导作用。

关键词 数据可视化,专业分流数据,双层放射环,学生管理
中图法分类号 TP391.9 **文献标识码** A

Double Sunburst Matrix Visualization to Overview Majors Distributary Data

LI Hui[1] CHEN Hong-qian[2] DONG Shuang[1] MA Li-yi[1]
(College of Management, Beijing Union University, Beijing 100101, China)[1]
(School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China)[2]

Abstract To show and analyze the feature of attribute flow in majors distributary data, a sunburst matrix visualization method was proposed in the paper. The data with various attributes are firstly selected and counted in the method. The statistical data are mapped to the visual elements of visualization results. The mapping procedure includes three parts. In the first part, the scattered bubble charts are introduced to express the whole statistical numbers of student with various source and destination. Secondly, contrast method of pie chart is introduced into bubbles to show its gender proportion. The pie chart is adopted as the inner layer of sunburst. Thirdly, the students' grade point attribute in each category are designed to display comparably in the outer layer of sunburst. The experimental results and the evaluation of the management staff denoted that the visualization results can be expressed directly to the flow characteristics such as the number of students, gender proportion, and grade point and so on. The detailed classification of students and professional construction and training are expected to achieve according to the visualization results.

Keywords Data visualization, Majors distributary data, Double sunburst, Management of student

1 引言

随着高等学校教育体制改革的不断深化,我国高校的人才培养经历了从"进校即分流"的单纯专业教育到"进校后分流"的通识教育与专业教育相结合的发展历程。目前,国内本科阶段教育中普遍实行"大类招生"。"大类招生"侧重于知识的横向联系,能够有效促进学科交叉和思维启迪,强调因材施教和科研创新[1],其目的是加强通识教育,适应社会对人才的需求,培养具有拓展知识领域的潜能且具备终身学习能力的人。学生进入大学经过一年多的学习后,在对所在的专业有一定了解的基础上再进一步选择更小的专业方向,就是所谓的专业分流。

对于专业分流成效的评估,目前主要有两类分析方法:1)通过调查问卷形式获取数据,然后通过问卷数据分析的方法判断分流成效;2)针对分流过程中所产生的数据,使用 SPSS 等统计分析软件得出分析结论,如王卫红[2]应用 SPSS 对专业分流中学生选报志愿、分流后各专业的学生绩点、分流前后课程成绩等进行分析和改进。

在专业分流数据分析过程中,主要希望获取如下 3 方面信息:1)希望获知某个专业的学生的原班级分布情况,这可以直接考查学生的分类志愿是否受班主任宣讲的影响;2)希望获知在专业分流过程中各班级流动到某专业中的学生绩点的分布情况,这将直接影响到该分流后专业的建设决策;3)希望获知专业分流是否出现性别比例失衡的情况。

本文针对专业分流过程中的学生志愿、分流过程、分流结果,以及学生性别、绩点等属性数据,提出一种名为双层放射

本文受北京市自然基金资助项目:针对影视数据的个性化推荐关键技术研究(9164028)资助。
李 慧(1983—),女,博士,讲师,主要研究方向为数据挖掘与可视化,E-mail:lihuimuzi@126.com;陈红倩(1982—),男,博士,副教授,硕士生导师,主要研究方向为数据分析与可视化;董 爽(1969—),女,博士,副教授,主要研究方向为数据挖掘与企业信息化,E-mail:gltdongshuang@buu.edu.cn(通信作者);马丽仪(1977—),女,博士,主要研究方向为商务智能。

环矩阵的可视化方法。该方法首先使用矩阵表示学生分流前后的班级，然后使用双层放射图表示每一个分区的学生性别比例和绩点分布情况，从而获得整个分流中的学生及其属性的流动情况的概览性可视化结果，使学生管理人员能够从可视化图形中一目了然地了解该次专业分流中学生的分布情况，以及分流过程中学生属性的分布情况。对该系统的使用结果表明，文中所提方法能够对专业分流数据进行多属性概览展示，实现多重属性下的分流数据对比。

2 相关工作

专业分流是按大类招生培养的重要环节，能否达到预期效果，直接关系到大类招生人才培养模式的成功实行与否。良好成效的专业分流，不仅可以促进各学科均衡发展，还能提升学生的专业满意度，提高学生的学习兴趣和学习积极性。颜兵兵[3]基于成绩优先和志愿优先两个原则提出了两种专业分流的实施细则，并通过实例进行比较分析，挖掘大类招生模式下专业分流过程中的表象与内涵。

针对专业分流数据的分析，能够深入了解学生专业分流新班级的来源结构、专业满意度以及专业兴趣度、教学进度的适应度等。对专业分流过程中出现的问题，要有充分的预见，并采取积极的对策；对专业培养方案、未来发展趋势以及推动各学科和专业的均衡发展方面，也应具备一定的借鉴作用。

可视化可以有效提高数据分析的效率，已经被证明为一种提高信息获取能力的有效方法[4]。目前国内外许多大学和研究机构都致力于可视化技术理论及应用的研究，如美国国家宇航局埃姆斯研究中心、奥地利虚拟现实与可视化研究中心、德国斯图加特大学可视化与交互系统研究小组、清华大学、浙江大学 CAD&CG 国家重点实验室等。

在可视化方法的实现原理方面，Neumann[5]给出了一个可视化信息分析技术的框架。根据可视化方法的实现原理，可视化技术主要分为基于几何的可视化技术、面向像素的可视化技术和基于图标的可视化技术。Keim[6]同时考虑数据类型、表现功能和交互功能 3 个方面，增加了标准 2D/3D 显示和重叠显示，并将文本/Web 页、算法/软件作为可视化的数据类型。

针对专业分流数据的分析，主要从数据中找寻属性间的相关性关系。对于相关关系的可视化，散点矩阵、气泡图都是常用的可视化方法。徐沛原[7]使用层次气泡图对基于移动应用程序日志数据的人群特征进行分析与画像计算，将人群特征与时间信息、地理位置信息结合，从多角度可视化地展现人群特征。Zhao[8]在交互过程中针对可视化对象进行动态关联，挖掘并分析信息之间的显式和隐含语义关联关系。

浙江大学的陈为教授团队[9]于 2013 年从不确定性的可视表达形式角度出发，梳理并总结了当前主流的图标法、可视变量编码法、几何体表达法和动画表达法等可视化方法。

对于多维数据可视化来说，平行坐标技术[10]和散点图矩阵技术[11]是常用的可视化方法，用于对多维数据间的关联性进行研究。Jorge[12]于 2014 年结合散点图投影技术和平行坐标，对不同数据内容间的相似度进行了可视化。Hofmann[13]针对多维数据关联可视化中的视觉错觉，提出了纠正该错觉的有效方法。

集成多种可视化方法的优点，将尽可能多的数据属性内容映射至同一可视化窗口甚至是同一可视化元素，是提高可视化能力的一种重要方法。文献[14]使用多视图协同的方法将平行坐标、散点矩阵等方法同时展现在一个可视化窗口中。针对可视化过程中的逐步分析，Charles[15]提出了一种渐进式可视分析方法，其能够对整个数据集进行分步探索式可视化。

3 双层放射环矩阵可视化方法

3.1 可视化方法设计

本文针对专业分流过程中数据的多属性快速分析问题，提出一种放射环矩阵的可视化方法。该方法的设计思想主要由 3 个步骤完成：

(1)将所有参与专业分流的学生根据其原班级和选择专业进行分类筛选，进行统计后映射为气泡图，其中原班级映射为 X 坐标，选择专业映射为 Y 坐标，人数映射为气泡大小；

(2)将每个分类的学生根据其性别属性进行再次分类统计，将性别属性映射为饼图，并将该饼图作为单体放射环的内层；

(3)将每个分类的学生根据其绩点属性进行分类统计，将绩点属性映射为单体放射环的外层内容，其中各绩点区间的人数映射为外层环宽值，绩点区间的比例关系映射为外层各扇形区的角度比例。

本文所提出的可视化方法设计原理即数据与可视化元素的具体映射关系如图 1 所示。

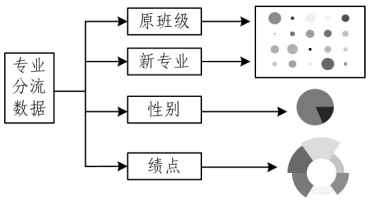

图 1 双层放射环设计原理图

本文所提出的可视化方法中的数据内容与可视化元素的映射关系如表 1 所列。

表 1 放射环内层的可视化元素的映射关系

专业分流数据	可视化元素
原班级信息	双层放射环所在 X 坐标
分流专业	双层放射环所在 Y 坐标
学生人数	放射环内层大小
性别比例	放射环内层扇区比例
不同绩点区间的学生人数	放射环外层扇区半径

3.2 矩阵布局方法

作为一种层次数据可视化方法，Sunburst 放射环也常用于多属性数据的可视化。本部分借鉴气泡图和放射环的形式，对专业分流过程中的学生原班级与专业选择关系进行可视化，以快速获得各个专业的学生来自于各班级的比例情况。

3.2.1 双层放射环最大半径计算

放射环矩阵是借鉴气泡图的设计思想，将 X 坐标、Y 坐标以及气泡大小均映射为数据属性，从而将数据展现在可视化结果中。其中每一个双层放射环单体在整个可视化结果中需要做到不互相覆盖的情况，因此需要首先根据最终可视化结果的可视区域计算双层放射环的最大半径，以充分利用可视面积。

首先根据最大高、宽计算出其所占矩形的范围，然后根据矩形计算其内接圆的半径最大值。双层放射环的最大半径的

计算方法如式(1)所示:

$$R_m=\max(R_1,R_2)$$
$$R_1=\min(\frac{H}{m},\frac{W}{n}) \tag{1}$$
$$R_2=\min(\frac{W}{m},\frac{H}{n})$$

其中,H 和 W 分别是可视区域的高度和宽度,m 和 n 分别是专业分流过程中的原班级数和分流专业数,根据数据分析需要,n 也可以作为新班级数。

3.2.2　双层放射环单体中心位置的计算

双层放射环的单体中心位置,需要根据不同的映射情况计算得出。首先需要根据可视区域、原班级数和分流专业数选择最佳映射状况,然后计算横、纵坐标位置。

对应(横坐标索引为 i,纵坐标索引为 j)的双层放射环单体中心位置的计算方法如式(2)所示:

$$x_i=M_L+i*inv_x$$
$$y_j=M_T+j*inv_y \tag{2}$$

其中,M_L 和 M_T 分别是在最终可视化效果中可视区域左侧和上侧的空白区域的宽度和高度;i 和 j 分别是双层放射环单体所对应的横、纵坐标索引值,在本方法中可通过式(3)得出;inv_x 和 inv_y 是矩阵布局在横、纵方向上的间隔值,可根据式(4)计算得出。

$$\begin{cases} i=1,\cdots,n,j=1,\cdots,m, & R_1\geqslant R_2 \\ i=1,\cdots,m,j=1,\cdots,n, & R_1<R_2 \end{cases} \tag{3}$$

$$inv_x=\begin{cases} W/n & R_1\geqslant R_2 \\ W/m & R_1<R_2 \end{cases}$$
$$inv_y=\begin{cases} H/m & R_1\geqslant R_2 \\ H/n & R_1<R_2 \end{cases} \tag{4}$$

式(3)和式(4)中的 H,W,m,n,R_1 和 R_2 的意义与式(1)中的意义相同。

3.3　放射环单体计算方法

每个双层放射环单体由内层和外层两层可视化元素组成,其中内层大小映射为该分类下的学生总人数以及学生的性别比例,外层映射为该分类下的学生成绩绩点情况。每个双层放射环单体的可视化设计原理如图 2 所示。

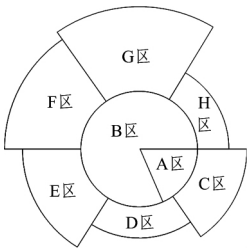

图 2　双层放射环单体的设计原理图

在该双层放射环单体中,A 区和 B 区组成内层环,其半径映射该分类下的学生总数,其中 A 区映射为性别为女,B 区映射为性别为男;C 区－H 区分别映射绩点为无绩点、0-1、1-2、2-3、3-4、4-5,一共 6 个区间,每个扇区的半径映射为该区间内的学生人数。

3.3.1　各放射环的内层半径的计算方法

双层放射环单体中,需要将根据式(1)所得的最大半径 R 分配 2/3 作为内层半径的最大值。第 i 个放射环的内层半径的计算方法如式(5)所示:

$$r_i=\begin{cases} 0, & NC_i=0 \\ \dfrac{2*NC_i*(R_m-R_0)}{3*NC_{\max}}+R_0, & NC_i>0 \end{cases} \tag{5}$$

其中,$NC_{\max}$ 是所有分类的最多学生数;R_m 是式(1)中计算得出的最大半径值;R_0 是系统设定的一旦值不为 0 时的最小映射值,该设定是为了保障小数值对应的汽泡的可见性;NC_i 是该分类下的学生数。

3.3.2　各放射环的外层半径的计算方法

双层放射环单体中,需要将根据式(1)所得的最大半径 R 分配 1/3 作为外层半径的最大值。第 i 个放射环中外层的第 j 个扇区半径的计算方法如式(6)所示:

$$R_{i,j}=\begin{cases} r_i, & NS_{i,j}=0 \\ \dfrac{NS_{i,j}*(R_m-R_0)}{3*NS_{\max}}+R_0+r_i, & NS_{i,j}>0 \end{cases} \tag{6}$$

其中,$NS_{\max}$ 是所有分类的所有绩点区间的最多学生数,R_m 和 R_0 与式(1)中的意义相同,$NS_{i,j}$ 是该分类下该绩点区间的学生数。

3.3.3　内外层扇区角度比例的计算

在每个双层放射环中,内层环根据男女生个数比例划分为两个扇区;外层环根据 6 个绩点区间平均划分为 6 个扇区,每个扇区的角度为 60°。

3.4　可视化元素着色

为提高数值区分度,在本文所提出的可视化方法中,将双层放射环的各区域进行着色。分别将性别和绩点区间进行颜色映射,颜色的映射值如表 2 所列。

表 2　双层放射环颜色映射值

数据值/区间	颜色名称	颜色 RGB
性别(无效值)	LightSteelBlue	176,196,222
性别(女)	MediumBlue	0,0,205
性别(男)	Magenta	255,0,255
绩点(无效值)	SandyBrown	244,164,96
绩点[0,1)	LightSlateGray	119,136,153
绩点[1,2)	LightSkyBlue	135,206,250
绩点[2,3)	Aqua	0,255,255
绩点[3,4)	Lime	0,255,0
绩点[4,5]	OrangeRed	255,69,0

4　实验结果与分析

将本文提出的针对学生专业分流数据的双层放射环矩阵可视化方法用于实际案例的数据分析,以检验该可视化方法的有效性。

所采用的专业分流数据是某高校连续 3 年的专业分流数据,该数据包含 1234 名同学的性别、绩点、分流前的班级、分流后的专业和班级等信息。

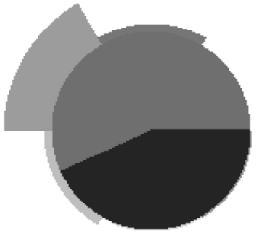

图 3　双层放射环单体可视化结果图

图 3 为专业分流过程中某个班级分流到某专业的双层放射环单体效果图。从该单体图中能够得出如下结论:

1)在该分类学生中,女生比男生多,女生比例约为 60%;

2)在该分类学生中,学生的绩点大致有 3 类:[1,2),[2,3),[3,4),其中绩点在[2,3)区间的人数最多。

图 4 是针对案例数据集的可视化效果图,当所有放射环单体共同组成双层放射环矩阵时,将能够得出进一步的结论。从该效果图中能够得出结论:

1)原班级 1、原班级 2 在专业分流过程中存在非常明显的倾向性,其中原班级 1 的学生绝大部分选择了专业 D,而原班级 2 的学生绝大部分选择了专业 C,原班级 3 和原班级 5 也存在较为明显的专业倾向性,这可能与分流前的原班级班主任或任课教师有关系。基于该结论可以进一步深入调查原因。

2)原班级 1,2,10 这 3 个原班级的学生的绩点普遍较高,而其他班级的绩点水平较低,尤其原班级 5,7,8 中绩点超过 2 的同学较少。

3)专业 C 和专业 D 从各个原班级中吸引到了绩点较高的学生,各个班级的绩点为[2,3)区间内的同学几乎被吸引到该专业中;而与此相反的是,专业 A 和专业 B 虽然吸引的学生总数也较多,但学生的绩点大多在[1,2)区间。这说明专业 C 与专业 D 对学生有更大的吸引力。

4)该数据集中的学生,总体而言女生较多而男生较少,分流后在专业 C 的女生比例更大,男女生比例更为失衡;而专业 E 和专业 B 的男女生比例相对较平衡。

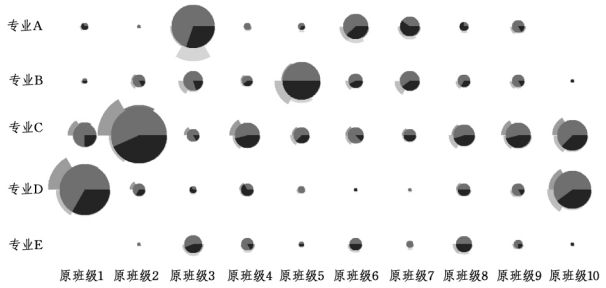

图 4 双层放射环矩阵的可视化结果

与已有方法相比,本文所提出的方法用于专业分流数据时能体现出更好的直观性与有效性。本文所提出的可视化方法能对分流过程中的学生人数、性别、绩点等属性的流向特征进行直观表达,并能够对专业分流中的学生进行快速的分类与细化,对专业建设与培养计划的制定起到良好的指导作用。

在可视化方法的性能分析方面,针对本文提出的可视方法的分析包含两部分:1)可视化方法的时间效率;2)可视化方法能同时展现的数据容量。在可视化方法的时间效率方面,可视化方法的计算复杂度主要体现在数据筛选和数据统计方面,这取决于所采用的数据库的时间效率,在实验过程中发现,专业分流数据的规模一般为几百学生,能够完全支持实时可视化;在可视化方法的数据容量方面,取决于数据筛选条件的值域范围,实验过程中发现,当横、纵坐标对应的条件值小于或等于 20 种时可视化效果最佳。

结束语 本文为分析专业分流过程中的学生属性特征和分布规律情况,设计了一种双层放射环矩阵的可视化方法,通过可视化元素展现数据,从而发现数据中的规律,并进行数据对比分析。该方法能够对各专业和班级学生的流动关系进行直观、快速的概览,并对各类学生中的性别属性和成绩绩点属性的分布规律和分类特性进行快速分析。

案例数据的可视化结果表明,本文提出的可视化方法能够提高学生管理人员的分析和管理效率,对专业分流后的针对性学生培养和专业建设起到良好的辅助作用。

在可视化方法的展望方面,本文提出的可视化方法属于多种数据属性内容的可视化方法,除能应用于专业分流数据外,还能用于各类个体数据的条件对比,如符合哪些条件的个体统计的可视化。

参 考 文 献

[1] 董泽芳,王晓辉. 普林斯顿大学本科人才培养模式的特点及启示[J]. 高教探索,2014(2):77-81.

[2] 王卫红,蔡雪. 计算机大类专业分流的实践与思考[J]. 计算机教育,2014(6):87-90

[3] 颜兵兵,姜永成,殷宝麟. 大类招生模式下专业分流方案设计与实例分析[J]. 高教论坛,2014(9):44-47.

[4] 袁晓如,张昕,肖何,等. 可视化研究前沿及展望[J]. 科研信息化技术与应用,2011,2(4):3-13.

[5] NEUMANN P,TANG A,CARPENDALE S. A Framework for Visual Information Analysis:Technical Report 2007-87123[R]. University of Calgary,Calgary,AB,Canada,July 2007.

[6] KEIM D,QU H,MA K L. Big-Data visualization [J]. Computer Graphics and Applications,IEEE,2013,33(4):20-21.

[7] 沛原. 基于移动终端日志数据的人群特征可视化研究[D]. 杭州:浙江大学,2015.

[8] ZHAO J,COLLINS C,CHEVALIER F,et al. Interactive exploration of implicit and explicit relations in facet datasets[J]. IEEE Trans. on Visualization and Computer Graphics,2013,19(12):2080-2089.

[9] 雷辉,陈海东,徐佳逸,等. 不确定性可视化综述[J]. 计算机辅助设计与图形学学报,2013,25(3):294-303.

[10] 孙扬,封孝生,唐九阳,等. 多维可视化技术综述[J]. 计算机科学,2008,35(11):1-7.

[11] ELMQVIST N,DRAGICEVIC P,FEKETE J D. Rolling the dice:Multi-dimensional visual exploration using scatter plot matrix navigation [J]. IEEE Trans. Vis. Comput. Graph. ,2008,14(6):1539-1148.

[12] POCO J,DASGUPTA A,WEI Y X,et al. SimilarityExplorer:A Visual Inter-Comparison Tool for Multifaceted Climate Data [J]. Eurographics Conference on Visualization (EuroVis),2014,33(3):341-350.

[13] HOFMANN H,VENDETTUOLI M. Common Angle Plots as Perception-True Visualizations of Categorical Associations [J]. IEEE Transactions on Visualization and Computer Graphics,IEEE Computer Society,2013,19(12):2297-2305.

[14] 陈谊,蔡进峰,石耀斌,等. 基于平行坐标的多视图协同可视分析方法[J]. 系统仿真学报,2013,25(1):81-86.

[15] STOLPER C D,PERER A,et al. Progressive Visual Analytics:User-Driven Visual Exploration of In-Progress Analytics[J]. IEEE Transactions on Visualization and Computer Graphics,2014,20(12):1653.