

doi: 10.7690/bgzd.2024.11.008

一种基于层次聚类的人力数据标签体系设计与实现

明丽洪, 马永一, 罗昌俊, 付渲理

(中国空气动力研究与发展中心计算空气动力研究所, 四川 绵阳 621000)

摘要: 人力数据作为各业务领域数据的核心承载点, 是连接人、物理实体、业务活动的关键点。通过设计人力数据标签体系, 将人力数据属性抽象为图结构, 通过层次聚类算法实现不同属性集的自动聚类 and 人才梯队划分, 可按照专业领域方向、科研项目承担、科研成果获得、人才奖励等维度实现人力数据的抽取和灵活组合呈现, 为人才梯队层级的培养规划与目标拟定奠定基础。

关键词: 人力资源; 数据标签; 层次聚类

中图分类号: TP311.13 **文献标志码:** A

Design and Implementation of Human Data Label System Based on Hierarchical Clustering

Ming Lihong, Ma Yongyi, Luo Changjun, Fu Xuanli

(Institute of Computational Aerodynamics, China Aerodynamics Research and Development Center, Mianyang 621000, China)

Abstract: As the core bearing point of data in various business fields, human resource data is the key point to connect people, physical entities and business activities. By designing the label system of human data, abstracting the attributes of human data into graph structure, and realizing the automatic clustering of different attribute sets and the division of talent echelon through hierarchical clustering algorithm, the extraction and flexible combination of human data can be realized according to the dimensions of professional field direction, scientific research project undertaking, scientific research achievement acquisition and talent award. Lay the foundation for the training planning and target formulation of the talent echelon level.

Keywords: human resources; data label; hierarchical clustering

0 引言

随着信息技术的快速发展,数据的使用已从“自生自用”“自生他用”逐渐演化为“共生共用”的模式。在这个万物互联的时代,数据成为了人类文明史上继蒸汽和电力之后的第三种重要能源,是人类进入数字文明时代的主要推动力,具有改变社会、改变未来的深远意义。谁能掌握数据,谁能灵活用好数据,发现数据间潜在的规律,谁就能在这个“未来已来,一切重构”的时代中立足。

人力资源数据作为各业务领域数据的核心承载点,可以与事物发生关联,也可以与人发生关联,是连结万物、发挥人的主观能动性、评价人的效能的重要关键节点^[1-2]。因受制于各个阶段信息化的发展,在当下以满足业务自身管理需求为主线构建的信息系统较多,大多数未考虑数据综合应用场景,导致数据孤岛、烟囱林立现象较为普遍,数据链路未打通,不同系统间对人力数据、角色数据、权限管理数据重复构建也较多,且标准不统一,给管好

数据、用好数据带来极大的困难。

以某科研院所信息化发展为例,目前已在其科研、试验、管理等领域建设了大大小小的信息系统上百个,一个系统一套数据,数据之间彼此相互割裂,导致数据重复冗余、数据治理与共享困难,数据再利用更是遥不可及,更无法谈及决策智能。所谓决策智能就是将数据科学、决策科学、社会科学和管理科学汇集到一个统一的领域,帮助人们使用数据来改善他们的生活、业务和世界。其中,决策科学更加关注根据可用信息做出最佳选择,数据科学更加关注使用科学的方法、流程、算法和系统从嘈杂的结构化及非结构化数据中提取知识和见解,并将知识和可操作的见解从数据中应用到广泛的应用领域。

在实践数字化转型落地中,首当其冲的就是要考虑数据和打通各类业务系统间的数据链,使得数据信息共享变成可能,为数据共生共用创造基本条件^[3]。人力资源数据作为以人为核心的信息资源

收稿日期: 2024-06-19; 修回日期: 2024-07-22

第一作者: 明丽洪(1979—),女,云南人,硕士。

基础，是打通各业务活动域信息系统的纽带与桥梁^[3-5]。

笔者结合某研究院所的发展现状与特征，提出了一种基于层次聚类的人力数据标签体系设计与实现方法，为人力数据在各业务领域灵活流转、按需组合应用提供基础，并打通以人为核心的各业务数据链。

1 设计

1.1 设计思路

首先运用思维导图工具设计出人力标签类目体系结构，然后从不同的维度方向延伸细化对象属性；其次，根据人力标签类目体系结构图再进行细分延伸，形成具体的标签设计列表，并为每个标签进行编号，形成散点图；最后运用层次聚类方法^[6]，对散点图进行自动分类聚集形成人力资源标签社区和人才梯队层级。

1.2 要解决的问题

1) 对现实环境中的人力资源构成进行详细梳理；2) 设计人力标签类目体系结构，从不同的维度方向延伸对象的属性；3) 对标签类目和标签进行编号；4) 将标签类目和标签抽象为图结构；5) 运用层次聚类算法，对标签进行自动聚类，形成人力资源标签社区和人才梯队层级。

2 实现方法

2.1 对现实环境中的人力资源构成进行详细梳理

对现实环境中的人力资源构成进行详细梳理，确定对象是抓住问题本质的关键。按照为“人”“物”“关系”进行数字映射，将现实世界的事物归属为 3 大类型对象。人作为发起行为的主体，具有主动性，需要多个维度进行分类梳理；物是行为中被施与的对象，往往是被动的，以业务维度进行分类梳理；关系则是人和人、人和物、物和物在某时刻发生的某种连接，是两两实体间的联系与连接，属于虚拟对象，它包括行为关系、归属关系、思维关系等各种强、弱关系，主要以业务流程为纽带进行强

关联。某科研院所数据对象如图 1 所示，从科研院所真实环境中识别出各类对象，并建立关系。例如，以人为核心的对象按照管理维度可分为管理者、科技人员和协作单位人员，以物为核心的对象按业务维度可分为科学研究项目、科学试验项目、科研研究设备、科学试验设备，以及人与物基础上建立的关系，其中包括科学研究、科学试验、科研设备采购、科研设备建设、科研设备运维及办公等。

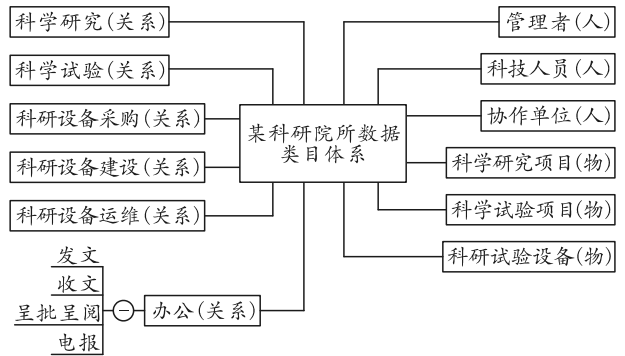


图 1 某科研院所数据对象

按照 3 个维度进行梳理，分别是以“人”的维度组织的数据、以“物”的维度组织的数据和以“流程”的维度组织的数据(即人与人，人与物及物与物发生的关系所产生的数据)。

2.2 设计人力标签类目和标签并编号

按“人”的维度组织的数据，可划分为管理者、科技人员和外协单位人员，管理者又进一步细分为行政管理者 and 业务管理者，科技人员按职称维度又可分为研究员、副研究员、助理研究员、正高级工程师、高级工程师、工程师、助理工程师、高级讲师、讲师、助教等，又可以按人员类型维度分为特种类型 1、特种类型 2、特种类型 3、职工、聘用等。按“人”的维度组织的数据都有其共性特征属性，将这部分共性特征属性抽取出来可进一步划分为基本属性、工作属性、教育属性、履历属性、社会关系属性、住房属性、车辆属性、健康属性、科研属性等，在上述大类属性类目下沉梳理至各个对象所包含的属性集，即可以进一步细分至字段级的属性。人力标签属性分类如图 2 所示。

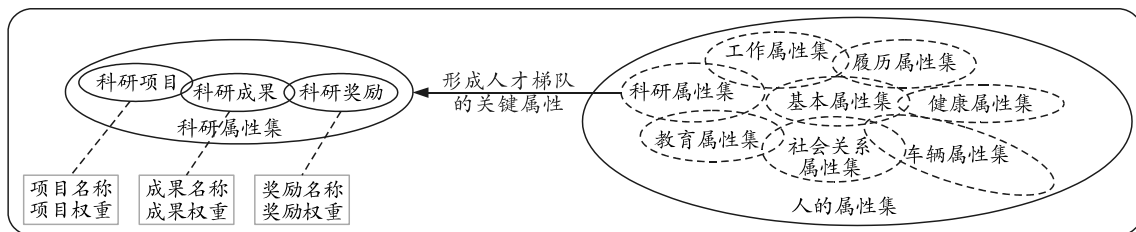


图 2 人力标签属性分类

2.3 将标签类别和标签抽象为图结构

将人力资源属性(即标签)抽象为一个网络图 $G(V, E, W)^{[7]}$, 其中 V 表示节点集, E 表示边集, W 表示链路权重集。 $W = \{w_{ij} | w_{ij} = f_c(i, j)\}$, 是由每个边的链路权重 w_{ij} 构成, 链路 (i, j) 的权重是通过代价函数 f_c 计算得出。基于层次聚类的人力标签抽象结构如图 3 所示^[7]。

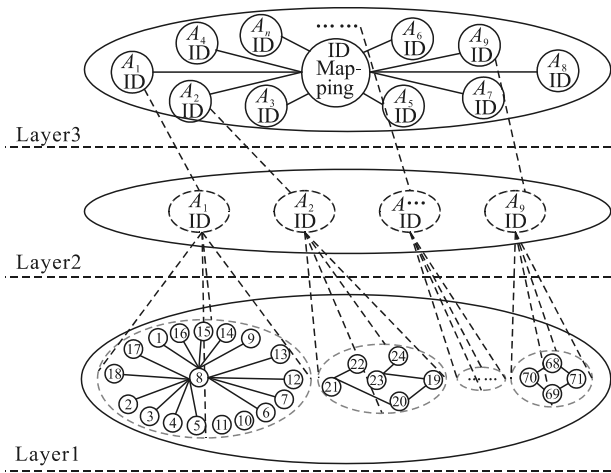


图 3 基于层次聚类的人力标签抽象结构

2.4 运用层次聚类算法, 对标签进行自动聚类

人才梯队聚类如图 4 所示, 自动聚类流程如图 5^[7]所示, 具体操作步骤如下:

1) 初始化: 将人力标签抽象为一个无向图, 并将所有节点作为子网划分的候选节点集, 计算图中各个节点的度, 并将节点按度的值从大到小排序。节点的度定义为^[6]: $Degree(i)$, 当节点 i 和节点 k 之间有链路存在时, $d(i, k) = 1$, 否则 $d(i, k) = 0$ 。

$$Degree(i) = \sum_{k=1}^N d(i, k) \quad (1)$$

2) 通过拓扑结构, 构建邻接矩阵。

3) 通过邻接矩阵及图中的链路、节点等基本信息, 求出网络节点间的相似度。相似度用于度量 2 个节点对之间连接的紧密程度。将图中节点 i 的邻接节点集合定义为 $Neibor_i$, 并且 $i \in Neibor_i$, 用数学表达为:

$$Neibor_i = \{j | A_{ij} = 1\} \cup \{i\} \quad (2)$$

式中 $(i, j) \in E$ 。共享邻接矩阵定义为 S , 矩阵各元素定义为 S_{ij} (即相似度):

$$S_{ij} = Neibor_i \otimes Neibor_j = \sum A_{kw};$$

$$k \in Neibor_i, w \in Neibor_j, \text{ 且 } (k, w) \in E \quad (3)$$

4) 根据关联度从弱到强, 逐步删除节点对之间的边, 根据需求设定停止删除的条件。

5) 得到的每个连通子图构成分裂的子集。

6) 从候选节点集中选择第一个节点(即度数最大的节点)及其所在的子集作为基本子网, 将这些点从候选节点集中移除, 并在此基础子网上通过改进的 d -ball 算法对该子网进行扩展。

7) 在候选节点集中选择与子网中现有节点直接相连的节点, 判断各个节点与该子网内的节点直接相连的边数是否大于与该子网外直接相连的边数, 若大于, 则将相应的节点加入该子网, 并从候选节点集中移除; 否则, 该节点不被划入当前子网。重复此操作, 直到不再有相应的节点加入当前子网, 或者子网节点数达到规定的上限值。

8) 重复 6)、7), 直到所有候选节点都被划入相应的子网。

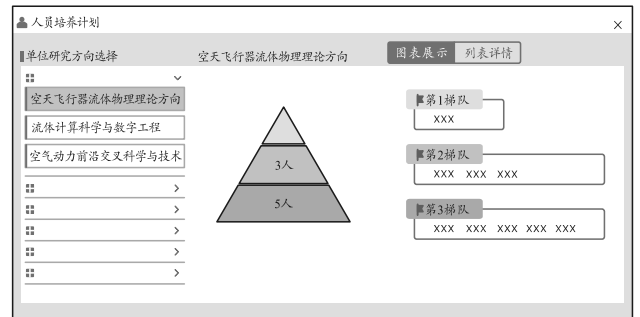


图 4 人才梯队自动聚类效果

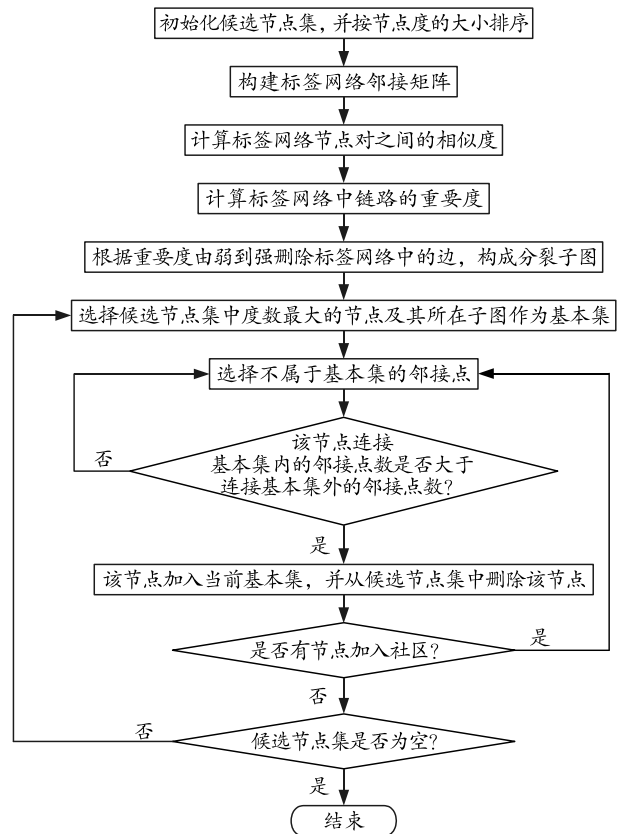


图 5 层次聚类算法执行流程

3 结束语

现实世界中的诸多系统均以网络形式存在。笔者运用同簇节点相互密集连接、异簇节点相互稀疏连接的网络簇结构特点，采用层次聚类算法对人力资源数据属性进行聚类划分，并综合不同业务领域数据分析形成不同的专业方向和人才梯队，为深挖人才潜力，培养规划专业团队方向，构建可持续的人才发展战略奠定决策基础。

参考文献：

[1] 张丽. 企业数字化人力资源管理转型的未来发作为[J]. 人力资源, 2021(10): 24-25.

(上接第 5 页)

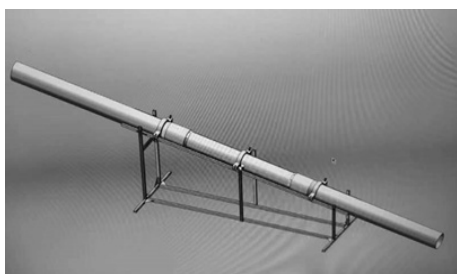


图 2 发射管

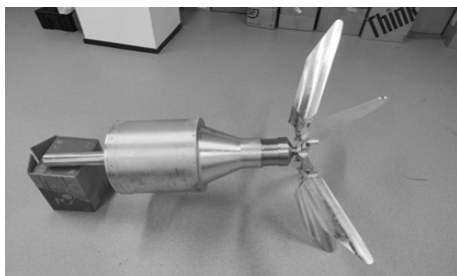


图 3 弹丸

试验结果表明：发射过载为 $6.6\sim 6.9\text{ m/s}^2$ ，弹丸飞行时间为 $2.4\sim 2.5\text{ s}$ ，安执机构全部解保。

(上接第 31 页)

5 结束语

笔者介绍了南网总调基于 DCCS 开发的调度指挥操作模块的功能设计及应用情况。为了提高调度操作效率，总调在该系统上设计开发了调度操作指挥模块，对命令操作、配合操作、许可操作和委托操作进行了设计研发，实现了调度操作全流程的网络化、信息化与智能化。模块投入使用后，数据显示对调度操作效率的提升作用显著。

[2] 纪伟国. 人力资源数字化的五大趋势[J]. 人力资源, 2021(1): 61-65.

[3] 梁云. 探究人力资源管理的数字化转型[J]. 人力资源, 2020(20): 28-29.

[4] 熊敏. 大数据时代企业人力资源管理变革的思考分析[J]. 人力资源, 2020(22): 148-149.

[5] 李温蔓. 数字赋能人力资源管理[J]. 人力资源, 2020(23): 108-110.

[6] 杨博. 复杂网络聚类方法[J]. 软件学报, 2009, 20(1): 54-66.

[7] 付渲理, 明丽洪, 李佳鹏, 等. 一种基于层次聚类的数据标签体系构建方法及系统: 中国, 1069209. 9[P]. 2023-08-24.

4 结束语

经理论计算和实际应用试验结果可知：该低过载安执机构解保试验方法是可行的，可为今后类似试验起到一定的参考，可为类似试验的一种标准方法。

参考文献：

[1] 周长省, 韩珺礼, 陈雄. 野战火箭发动机设计[M]. 北京: 国防工业出版社, 2015.

[2] 曹兵, 郭锐, 杜忠华. 弹药设计理论[M]. 北京: 北京理工大学出版社, 2016.

[3] 李二钢. 模拟导弹发射过程的导弹引信解保机构试验台——测试部分的研究[D]. 太原: 中北大学, 2011.

[4] 张艳兵, 武志博, 马铁华, 等. 引信保险机构双环境力加载模拟装置[J]. 探测与控制学报, 2017, 39(4): 22-25.

[5] 吕铁钢. 模拟制导炮弹引信环境的试验弹动态响应分析[D]. 太原: 中北大学, 2018.

[6] 武志博, 焦斌, 张艳兵, 等. 引信 MEMS 安解装置二维过载加载方法及仿真分析[J]. 火炮发射与控制学报, 2023, 44(6): 36-43.

参考文献

[1] 李群山, 李国统, 赖宏毅, 等. 基于状态令操作模式的一体化智能操作票系统研究与应用[J]. 智能电网, 2017, 5(2): 207-213.

[2] 周永灿, 邱生敏, 张坤, 等. 调度操作指挥网络发令“六步法”设计与实现[J]. 自动化仪表, 2022, 43(10): 33-38.

[3] 陈兴望, 辛阔, 孙雁斌, 等. 基于加权朴素贝叶斯算法的调度指挥态势感知模块设计[J]. 计算技术与自动化, 2022, 41(3): 121-127.

[4] 杨阳, 林泽宏, 曾远方. 调控一体化模式下电网调度管理优化研究[J]. 机电信息, 2021(15): 1-3.