

本文引用格式: 辛志杰. 基于模糊聚类的企业人力资源信息自动整合系统[J]. 自动化与信息工程, 2022, 43(5): 35-39.

XIN Zhijie. Enterprise Human Resource Information Automatic Integration System Based on Fuzzy Clustering[J]. Automation & Information Engineering, 2022, 43(5): 35-39.

基于模糊聚类的企业人力资源信息自动整合系统

辛志杰

(杨凌示范区人才交流服务中心, 陕西 杨凌 712100)

摘要: 人力资源信息管理是优化企业人力资源配置的基础。针对传统人力资源信息整合系统主要利用固定流程对数据进行统一配置处理, 忽略了人力资源信息中无明显分类界限的特点, 导致系统输出存在大量冗余, 整合效果较差的问题, 设计基于模糊聚类的企业人力资源信息自动整合系统。利用 Struts 框架搭建三层系统结构, 分析信息数据来源和对应实体关系, 建立系统数据库; 引入信息熵提取人力资源信息结构, 并使用模糊聚类算法聚类信息, 实现人力资源信息整合。本文系统与基于用户画像和基于知识元的整合系统进行对比, 其信息冗余比、整合差异比均最低, 信息整合效果好、效率高。

关键词: 模糊聚类; 企业人力资源; 信息整合; Struts 框架; 信息熵

中图分类号: TP311

文献标识码: A

文章编号: 1674-2605(2022)05-0007-05

DOI: 10.3969/j.issn.1674-2605.2022.05.007

Enterprise Human Resource Information Automatic Integration System Based on Fuzzy Clustering

XIN Zhijie

(Yangling Demonstration Zone Talent Exchange Service Center, Yangling 712100, China)

Abstract: Human resource information management is the basis for optimizing the allocation of human resources in enterprises. In view of the fact that traditional human resource information integration systems mainly use fixed processes to uniformly configure and process data, ignoring the characteristics of no obvious classification boundaries in human resource information, which leads to a lot of redundancy in system output and poor integration effect, an enterprise human resource information automatic integration system based on fuzzy clustering is designed. Use Struts framework to set up a three-tier system structure, analyze the source of information and data and the corresponding entity relationship, and establish the system database; Information entropy is introduced to extract the information structure of human resources, and fuzzy clustering algorithm is used to cluster information to achieve the integration of human resources information. Compared with the integration system based on user portrait and knowledge element, the system in this paper has the lowest integration difference ratio and information redundancy ratio, with good information integration effect and high efficiency.

Keywords: fuzzy clustering; enterprise human resources; information automatic integration; Struts framework; information entropy

0 引言

人力资源管理是企业经营的重要工作组成。企业经营规模扩大、业务量增加, 均提升了人力资源管理工作的难度。在企业信息化建设背景下, 日常经营管理也逐渐转为信息化^[1]。信息资源整合能够提升管理

工作效率, 对企业经营管理的重组与优化配置十分重要。信息资源整合的相关研究成果较多, 适用场景存在差异。文献[2]针对政务信息横向整合度低, 缺乏个性管理功能的问题, 利用用户画像技术, 通过分析大数据, 搭建信息整合模型。文献[3]利用知识元相关理

论, 设计在模型-视图-控制器 (model view controller, MVC) 软件架构模式下的信息整合系统, 提升数据利用与存储效率。文献[4]提出的整合系统利用信息检索和整合模块, 减少数据整合丢失量, 优化系统性能。

人力资源信息整合系统将企业人力资源管理过程中产生的信息数据进行组织配置, 拓宽人力资源信息的应用范围, 最大化利用人力资源, 增强人力资源管理能力。整合人力资源信息时, 由于资源信息数据与实体之间没有明显区分界限或属性, 导致传统的整合方法无法解决资源信息中的不确定性, 造成资源整合结果出现大量冗余。模糊聚类算法对数据的归属度划分不明确, 无需预先设定清晰的分类标准, 实际分类应用效果更佳。为此, 本文设计基于模糊聚类的企业人力资源信息自动整合系统, 满足人力资源管理工作的基础需求的同时, 进一步提高人力资源的使用与效率。

1 系统框架

基于模糊聚类的企业人力资源信息自动整合系统主要由显示层、业务逻辑层、数据层组成, 系统框架如图 1 所示。

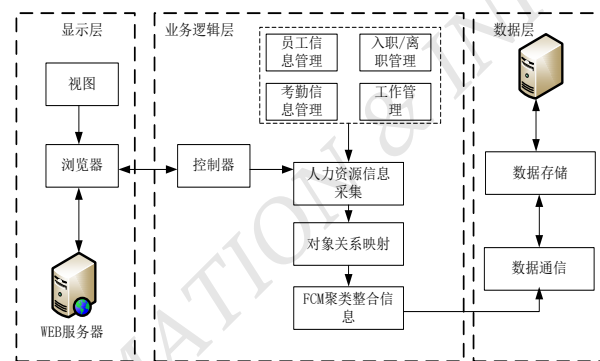


图 1 基于模糊聚类的人力资源信息自动整合系统框架

显示层主要用于人机交互。用户通过系统终端页面的 HTML 显示按键, 操作系统并获取响应。

业务逻辑层是系统的核心, 主要响应用户请求、采集、整合人力资源信息数据、返回指令请求等操作。

数据层对整合的人力资源信息数据进行分类存

储、调用处理。当业务逻辑层发出数据查询请求时, 数据层调用数据库的数据返回给业务逻辑层。

系统采用 Struts 框架搭建, 通过企业内网将人力资源管理部门的业务终端接入系统; 通过网络路由、防火墙将外部人力资源管理终端或子公司终端接入系统^[5]。在 B/S 架构下, 系统采用 HTTP 协议通信, 每次通信都需对数据请求接收形式进行封装、验证, 判断请求是否异常。

控制器接收资源信息管理终端的业务请求后, 发送给下一个系统组件; 经系统组件的映射、配置与请求对应的动作接口和路径, 执行业务处理逻辑操作。

2 数据库设计

为方便调用数据库内整合的人力资源信息, 本文在传统人力资源信息整合系统硬件的基础上, 利用 MySQL 对数据库进行优化设计。数据库中的数据以文本、表格、图片等文件形式存储。

系统的数据库可分为主题、业务、基础 3 个子数据库。其中, 主题数据库主要存储人力资源管理的不同操作信息, 包括人力资源数据修改、企业内部人力调动关系管理等; 业务数据库主要存储人力资源管理的相关数据, 包括各部门的人力信息、部门人力关系、人员和业务对应信息等; 基础数据库存储企业内部人力资源情况的基本数据, 主要包括考勤信息、在岗信息、每日工时等。主题数据库中操作实体与业务数据是一对多关系; 业务数据库中业务部门与人力数据之间为一对多关系; 人力资源信息数据与员工基础信息之间存在关联关系。

根据数据库中存储数据的实体关系, 以星型关系模型为架构, 利用 MySQL 2014 设计系统数据库。企业人力资源信息中主要涉及的实体包括员工、部门、在岗情况、对应业务等。根据实体与信息数据属性之间的关系, 建立如表 1 所示的数据库表。

用户登录系统, 通过显示层交互按键发出信息调用请求后, 数据库 GetUserInfo 类查找调用指令对应的数据表; 将数据表中的数据按照调用指令, 抽取对应的数据类信息数据; 数据经过 GetUserInfo 类的模

板文件渲染转换为对应显示层的显示形式，响应用户请求。

表1 系统数据库表（仅部分）

字段	数据类型	字段长度	是否为空	说明
ID	Int	12	空	员工 ID
Name	Varchar	20	空	员工姓名
S_info	Varchar	50	空	员工基本信息
Depart	Int	10	空	部门
Depart info	Varchar	20	—	部门基本信息
Jobinfo	Varchar	8	空	在岗信息
TIME	Char	18	空	信息更新时间

用户对整合的人力资源信息数据进行更新、删除等处理时，通过交互按键发出数据修改指令；数据修改指令通过 ModifyUserInfo 类转换为数据修改操作，并将数据修改结果返回给用户；修改结果在显示层显示，经用户确认后，由数据库存储。

3 模糊聚类算法整合信息

原始的人力资源信息经清洗、转换处理后生成无噪声数据。采用模糊聚类算法对人力资源信息进行整合，可减少为先验知识的依赖，提高信息整合效率。首先，读取待处理的原始人力资源信息，根据数据库的数据表设置种子节点；然后，以种子节点为参照，分割信息数据集；接着，信息数据集中的数字型数据若存在少量缺失，采用计算均值填补，若存在大量缺失，采用缺失值所处位置的中位数填补；最后，计算同类型信息之间的相似度。

$$\text{sim}(r_i, r_j) = \frac{\text{cov}(r_i, r_j)}{\sigma_i \sigma_j} \quad (1)$$

式中：

$\text{cov}(r_i, r_j)$ ——两个信息数据的协方差；

σ ——人力资源信息数据的方差。

考虑待整合人力资源信息的属性特点，系统设定

当 $\text{sim}(r_i, r_j) > 0.7$ 时，两个人力资源信息高度相似。

根据信息完整度高低，将高度相似信息中完整度较低的信息删除。

本文利用信息熵结构抽取原理，对人力资源信息进行全局抽取。根据人力资源信息特点，生成 M 个基础聚类，利用公式(2)计算基础聚类中存在相同数据点的聚类簇相关关系：

$$R_{ij} = \left| \frac{S_{c_i} \cap S_{c_j}}{S_{c_i} \cup S_{c_j}} \right|, i \neq j \quad (2)$$

式中：

S_c ——基础聚类簇中的人力资源信息样本数据。

利用公式(2)生成待处理的人力资源信息的一阶关系无向图。对不存在相同数据点的聚类簇，利用公式(3)计算二阶相似度。

$$\begin{cases} R'_{ijk} = \min \{ R_{ik}, R_{jk} \} \\ R'_{ij} = \sum_{k=1}^n R'_{ijk} \end{cases} \quad (3)$$

式中：

R'_{ijk} ——基础聚类中存在不相同数据点的聚类簇的相似度；

R_{ik} ——连续属性；

R_{jk} ——离散属性；

R'_{ij} ——基础聚类中存在相同数据点的聚类簇的相似度。

将聚类簇间的相似度与基础聚类簇的总量之比，作为各聚类簇间存在相关性的概率 $p(R'_{ij})$ ，计算人力资源信息整合信息熵。

$$H(R) = - \sum_{k=1}^n p(R'_{ij}) \lg [p(R'_{ij})] \quad (4)$$

由公式(4)得到的二阶相关性，建立人力资源信息结构矩阵：

$$S(c_i, c_j) = \begin{cases} 1, x_i = x_j \\ \frac{Z}{|q_{x_i}| |q_{x_j}|} H(R) \sum_{i=1, j=1}^N R_{ij}', \text{其他} \end{cases} \quad (5)$$

式中:

Z ——置信度;

q_{x_i} ——包含人力资源信息 x_i 的聚类簇集合;

q_{x_j} ——包含人力资源信息 x_j 的聚类簇集合。

提取人力资源信息结构后, 利用模糊 C 均值 (fuzzy C-means, FCM) 聚类算法聚类人力资源信息, 得到整合结果。

以信息熵提取的信息结构为初始聚类中心选择标准。若信息结构中共有 K 个不同的类, 则 K 个聚类的信息样本数据中心为 f_i , 信息数据 x_j 归属于对应聚类簇的模糊隶属度为 b_{ij} 。根据 FCM 聚类原理, 聚类的目标函数和约束条件为^[6]

$$\begin{cases} J(F, B) = \sum_{i=1}^K \sum_{j=1}^N b_{ij} \|x_j - f_i\|^2 \\ s.t. b_{ij} \in [0, 1] \\ \sum_{i=1}^K b_{ij} = 1 \end{cases} \quad (6)$$

利用 FCM 聚类算法分类人力资源信息时, 以拉格朗日乘子对公式(6)求极值, 得到单次聚类时的信息隶属度和单个聚类簇中心。选择新的聚类中心, 重复对公式(6)的极值求解过程, 直至无新的聚类中心产生, 停止模糊聚类过程。最终输出的模糊聚类结果, 即为企业人力资源信息整合的最终结果。

采用 JAVA 语言实现系统功能, 即完成系统设计。

4 系统性能测试

为验证系统性能, 对系统的信息整合性能进行测试。测试环境: 客户端配置, 4.9 GHz i7-12700 CPU, 64 GB 高频内存, 1 TB 高速固态内存 Windows 10 操作系统; 系统服务器配置, Intel Xeon 3.5 GHz 四核 CPU, 可扩展内存 128 GB, 数据库为 MySQL 2014。

4.1 测试指标

系统性能测试主要以整合后人力资源信息的冗余比和信息整合差异比为指标进行验证。

整合后人力资源信息的冗余比为

$$R = 1 - \left(\frac{Q}{mn} \right) \quad (7)$$

式中:

Q ——相邻属性值变化次数的累加和;

m ——行数;

n ——列数。

信息整合差异比为

$$D = \frac{A - E}{E} \times 100\%$$

式中:

A ——实际整合数据量;

E ——预期整合数据量。

4.2 测试数据分析

由基于模糊聚类的系统 (本文系统)、基于用户画像的系统 (系统 1)、基于知识元的系统 (系统 2) 分别对测试集信息进行整合, 整合后的信息冗余比如图 2 所示。

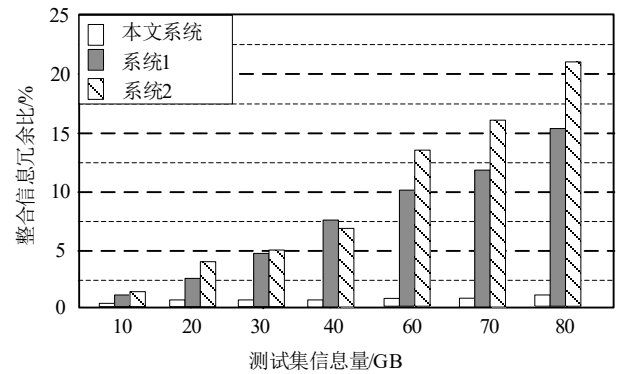


图 2 3 个系统整合后的信息冗余比

由图 2 可以看出, 基于模糊聚类的系统信息冗余比远低于另外 2 个系统, 且冗余量不完全随测试数据量增加而增加, 说明系统整合后不具价值的重复信息大量减少, 提升了后续的信息利用效率, 节

省存储空间。

在相同时间内,实际信息整合量与预期整合量的差异比如图3所示。

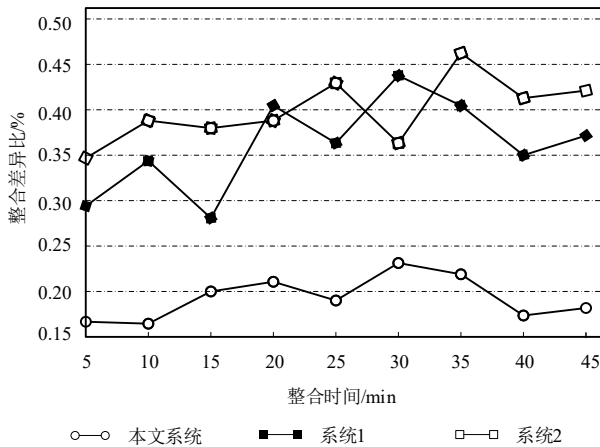


图3 实际信息整合量与预期整合量的差异比

由图3可知:基于模糊聚类的系统整合差异比低于另外2个系统,表明基于模糊聚类的系统整合信息量与预期更接近。从数值上看,基于模糊聚类的系统整合差异比小于0.23%,低于基于用户画像的系统最小差异比0.28%、基于知识元的系统最小差异比0.35%;同时,在相同时间内,基于模糊聚类的系统整合信息的速度明显高于另外2个系统。综上所述,本文设计的基于模糊聚类的企业人力资源信息自动整合系统信息整合效果好、效率高。

为进一步验证系统的实用性,以信息整合时长为指标,采用基于模糊聚类的系统(本文系统)、基于用户画像系统(系统1)与基于知识元的系统(系统2)分别进行测试,测试结果如表2所示。

由表2可知,基于模糊聚类系统的信息整合时长最长为1.40 min,基于用户画像系统的信息整合时长最长为3.99 min,基于知识元系统的信息整合时长最长为5.12 min。由此可见,基于模糊聚类系统的信息整合时长明显低于基于用户画像系统和基于知识元系统。基于模糊聚类的系统使用模糊聚类算法聚类信

息,解决了资源整合结果的大量冗余现象,提升系统的信息整合效率。

表2 3个系统信息整合时长对比表

测试次数/次	本文系统/min	系统1/min	系统2/min
10	1.05	3.16	4.89
20	1.15	3.11	5.01
30	1.32	3.67	4.95
40	1.27	3.95	4.68
50	1.37	3.65	4.99
60	1.26	3.99	5.03
70	1.34	3.65	5.12
80	1.40	3.21	4.36
90	1.39	3.87	4.85
100	1.38	3.21	4.36

5 结语

为实现人力资源信息有效利用,本文设计基于模糊聚类的企业人力资源信息自动整合系统。该系统利用模糊聚类算法,对不同特征的人力资源信息进行聚类,实现人力资源信息有组织地聚合处理,提升信息的利用价值。

参考文献

- [1] 方锦文,许潼歆,何晋乐.高校信息整合平台前端设计与实现[J].信息与电脑(理论版),2022,34(8):111-113.
- [2] 蔡盈芳,李子林,虞香群.基于企业用户画像的政务档案信息整合利用模型设计[J].档案学研究,2021(2):125-131.
- [3] 沙敏,刘广琦.MVC 模式下电子档案信息自动整合系统设计[J].现代电子技术,2020,43(22):90-93.
- [4] 梁琴琴,何东林,王振飞,等.三维数字化矿山地质信息整合系统设计与应用[J].中国金属通报,2021(1):237-238.
- [5] 梁凌,邓赵红,王士同.兼顾显隐信息与特征加权的多视角模糊聚类[J].计算机科学与探索,2021,15(6):1092-1102.
- [6] 毛俊华.基于模糊聚类的智慧医院多源异构数据整合系统[J].电子设计工程,2022,30(7):120-124.

作者简介:

辛志杰,男,1983年生,本科,副高级工程师,主要研究方向:信息管理与信息系统。E-mail: 17192984112@163.com