

多元回归模型在人力资源管理中的应用分析

李金霞

吉利学院

DOI:10.12238/ej.v7i8.1802

[摘要] 传统的人力资源管理已无法满足大数据时代背景下,企业对人才的需求。招聘过程中如何对海量的数据进行筛选和处理,为企业创造出更好的价值,选拔最合适的员工成为一个焦点。机器学习(ML)技术的发展为人力资源的管理提供了新的方向,注入了源源不断的动力和潜力。本文就是利用最优深度分类树的多元回归模型,首先通过分类树算法,确定树的最优深度,然后对招聘数据进行分析,确定信息的重要程度,再通过对均方误差(MSE)值的比较,建立起一种多元回归机器学习的模型。本模型不仅有效地提高人力资源管理在数据分析方面的效率,还为企业和员工之间如何做到最优的互选,提供了一定的借鉴。

[关键词] 大数据; 人力资源管理; 分类树; 回归模型

中图分类号: F062.1 **文献标识码:** A

Analysis of the application of the multiple regression model in human resource management

Jinxia Li

Geely University

[Abstract] Traditional human resource management has been unable to meet the demand of enterprises for talents in the background of big data era. In the recruitment process, how to screen and process the massive data, to create better value for the enterprise, and to select the most suitable employees to become a focus? The development of machine learning (ML) technology has provided a new direction for human resource management and injected a steady stream of power and potential. This paper uses the multivariate regression model of the optimal depth classification tree. Firstly, the optimal depth of the tree is determined through the classification tree algorithm, and then the recruitment data is analyzed to determine the importance of the information. Through the comparison of the mean square error (MSE) values, a model of multivariate regression machine learning is established. This model not only effectively improves the efficiency of human resource management in data analysis, but also provides some reference for how to achieve the optimal mutual selection between enterprises and employees.

[Key words] big data; human resource management; classification tree; regression model

引言

随着信息技术和教育事业的高速发展,为我国各个行业提供了众多的高素质人才,人力资源管理成为了企业和人才的纽带,起到了关键作用。这也给企业带了挑战和机遇,传统的人力资源管理职能已经满足不了企业的需求,带来的弊端也越来越明显,而人工智能(AI)和机器学习(ML)技术的发展为人力资源管理提供了新的方向,我们利用机器学习(ML),通过对数据的分析,可以为企业提供更好的人力资源解决方案,极大提高了人力资源管理工作的效率。

我们通常都是利用机器学习(machine learning)构建模型,

通过对测试数据分析,进行预测和决策。本文通过引入企业的招聘信息,主要包括:城市、公司规模、地区、教育层级、招聘专业要求、待遇、月薪、工作年限、发布招聘时间等重要信息,作为机器学习的样本数据,利用分类树的多元回归模型对企业招聘进行了一定的预测,这样不仅促进了企业人力资源管理的数字化转型,还提高了企业人力

资源管理水平^[2-3]。

1 机器学习与人力资源管理的研究现状

1.1 人力资源发展的概述

人力资源系统发展大致分成了三个阶段^[4]:

第一阶段: 人力资源管理系统的雏形, 他出现在20世纪60年代, 受到计算机软件和硬件的发展制约, 人力资源管理系统只是利用计算机进行简单的薪酬计算, 几乎不涉及管理和数据分析方面的问题。

第二阶段: 人力资源管理系统的形成, 出现在20世纪80年代初, 随着数据库技术的不断完善, 为人力资源管理系统提供了技术支持, 但此阶段也仅仅是以各类报表的形式呈现, 如: 财务报表和薪资数据报表, 仅限于实现数据的录入、查询和统计等基本功能, 但此时的系统还存在一些问题, 例如: 管理不够规范, 不能对数据进行系统和全面的分析, 信息的处理也不够全面等缺点。

第三阶段: 人力资源的完善阶段, 出现在20世纪90年代末, 由于社会和经济的快速发展, 企业的竞争非常激烈, 人才引进和人才培养成为了人力资源的重中之重, 由此也带来了人力资源要处理的数据急剧增加。21世纪后, 随着大数据时代的到来, 数据信息量呈几何倍数的增长, 我们如何从海量人力资源的数据中快速挑选出适合的数据, 如何有效地分析数据之间存在的关联, 预测企业未来的发展。机器学习(ML)技术的出现, 为人力资源提供了新的方向, 为人力资源管理的发展注入了源源不断的动力和潜力。

1. 2机器学习的研究现状

根据RS. Michalski等人的说法, 机器学习研究可以划分为“从样例中学习”, “在问题求解和规划中学习”, “通过观察和发现学习”和“从指令中学习”; 而根据E. A. Feigenbaum等人, 在《人工智能手册》中的划分, 机器学习可以分为“机械学习”, “示教学习”, “类比学习”和“归纳学习”^[1]。

伴随着大数据时代的到来, 机器学习已经被广泛的应用于人力资源管理领域,

它不但能有效的完成数据挖掘对大数据的信息提取和分析, 而且还能有效地分析数据之间存在的关联, 并对企业未来的发展做出一定的预测。早期关于数据挖掘方面的研究, 2008年, Ranjan J, Goyal D P, Ahson S I^[5]在论文中提出了人力资源管理系统中最好的决策: 数据挖掘技术。

2013年, 高超^[6]在论文中详细阐述了数据挖掘在人才管理系统中的重要作用。2016年, 潘现玮^[7]详细描述了数据挖掘在人力资源信息分析中的应用; 重点描述了决策过程中关于质量方面的问题; 2016年, 张金艳^[8]重点描述了数据挖掘在离职管理中的应用。2018年, 赖华强、王三银、仲崇高^[9]通过对人力资源管理系统离职管理的分析, 对数据挖掘的前景, 提出了具体的应用展望。数据挖掘与机器学习相结合的研究方面, 2008年, 刘宇^[10]提出了一种遗传算法的模糊聚类算法, 实现了对人力资源管理系统中的员工的绩效自动考核, 并且能够对算法提供的数据进行数据分析; 2012年, 姜飞^[11]提出一种利用机器学习, 通过对数据集的分析, 对岗位自动分配的算法, 实现了人力资源管理中岗位分配的功能。2016年, Huang M, Tsou Y, Lee S^[12]提出了模糊数据挖掘与模糊人工神经网络相结合的方法去预测员工的未来

业绩, 并且给出评估职位和项目分配的人员。

机器学习正逐步被应用于: 政府、教育、互联网、交通、电商、金融、医疗等相关领域, 因各个领域都涉及到人才, 我们现在被“数据丰富, 但信息贫乏”所困扰, 传统的人事管理系统往往在海量数据的筛选方面已经捉襟见肘, 更无法满足企业对大数据的需求, 故机器学习应用于人力资源管理已经迫在眉睫, 它可以通过建立自己的模型, 对大数据进行筛选处理, 最终将高数量级别的数据转换成低数据量级别的, 有用的数据和信息, 这样更有利于提高对人力资源的管理和对数据的分析。因此, 机器学习在人力资源管理中的应用具有非常重要的现实意义。

2 人力资源管理系统的简介

人力资源部门可以通过对大数据的分析和建立模型, 对战略决策, 企业的发展方向, 企业的发展趋势进行准确的预测, 尤其在人才招聘、培训、绩效等方面。随着大数据时代的到来, 人力资源管理部门发挥了越来越重要的作用, 我们可以通过机器学习对海量的数据进行深入的分析, 通过对关键性指标的透彻性解析, 帮助企业准确的预测员工的绩效、流失率、晋升潜力, 为企业提供根据具有前瞻性的策略, 以更好地应对和迎接未来的挑战。

企业人力资源管理系统主要包括: 人力资源的规划方面、招聘方面、培训、绩效管理、薪酬计算和离职处理等六大部分。

2. 1资源的规划方面

为了更有效的提高企业的效率, 确定企业的经营发展方向, 我们可以通过分析企业内外部的数据, 预测出更符合企业发展的决策和方向。

2. 2企业员工的招聘方面

企业传统的招聘方式是通过自己企业人力资源管理部门对招聘的人员进行多轮的面试, 最后确定录取的人员。在整个招聘过程中存在几个问题: 招聘信息的发布受到地域和时间的限制; 人工筛选简历的过程较为繁琐, 而且工作量巨大; 面试官的主观因素有时候对招聘的结果会产生巨大的负面影响; 而机器学习则会规避这些缺点。

2. 3企业员工的培训方面

随着知识技术的高速发展, 企业对新技术和新技能的需求不断增加, 人力资本势必会有所增加。如何进行培训, 怎么进行培训, 都培训什么内容将会成为一个难题? 而机器学习则会很轻松的解决这些问题, 它会针对每个员工给出个性化和最优化的培训计划。

2. 4在绩效管理方面

在企业的日常工作中, 企业资源管理部门主要是通过对员工的个人表现和业绩, 进行有效合理的评估; 而机器学习则是通过对员工的历史绩效、跟踪表现来预测员工的绩效, 我们可以通过对考评数据设置一定的权重, 不断更新和升级考核方案, 达到全方位收集数据和公平考核的目地。

2.5 薪酬方面

在薪酬方面,为了预测、调动和挖掘企业员工的潜力,我们可以通过机器学习的模型,不断的学习训练,自动设计和推演出最好的决策,已达到公平和客观的薪酬评估系统。

2.6 员工管理方面

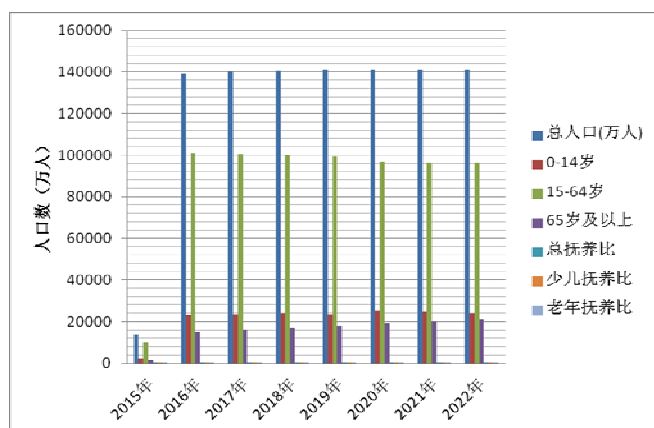
在员工管理方面我们主要通过忠诚度、满意度和敬业度三个指标来体现员工的工作关系质量,一般情况,我们都是通过问卷调研的形成来收集这三个重要的指标,但这往往过于主观,而且还受到很多客观因素的影响,如:工作环境、情绪、个人情感等客观因素的影响。而机器学习可以简单和快速的处理这些重复性的工作。

大多数机器学习应用领域只涉及到企业招聘和员工离职管理两个主要方面,本文就是针对企业招聘提出了一种最优深度分类树的多元回归模型,通过对收集的北京地区2019年400条招聘信息的数据进行分析,推断出招聘信息表中各个指标的重要程度,对招聘信息进行预测。

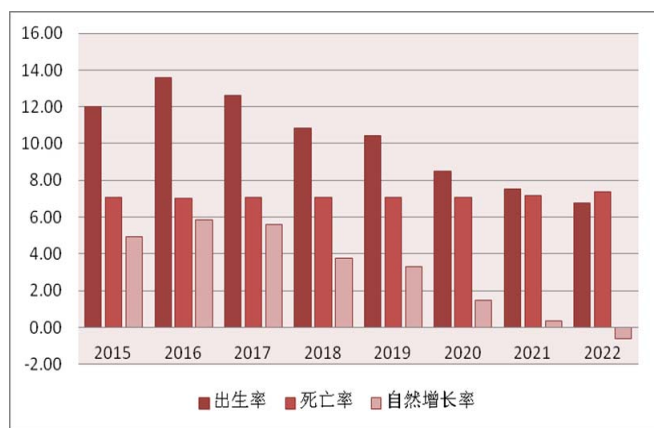
3 最优深度分类树的多元回归模型的建立

3.1 劳动力市场的分析

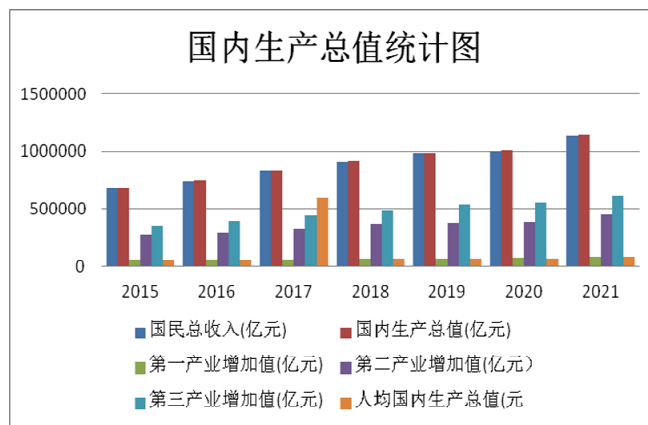
3.1.1 从国家统计局获取2015-2022年人口结构和抚养比,具体如下图所示:



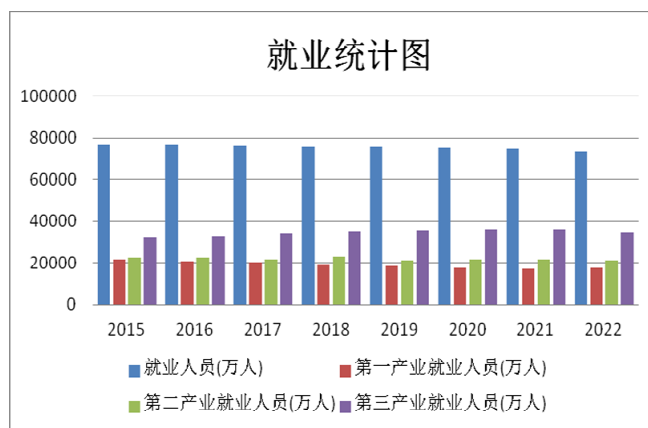
3.1.2 从国家统计局获取2015-2022年人口出生率、死亡率、增长率的比例图,具体如下图所示:



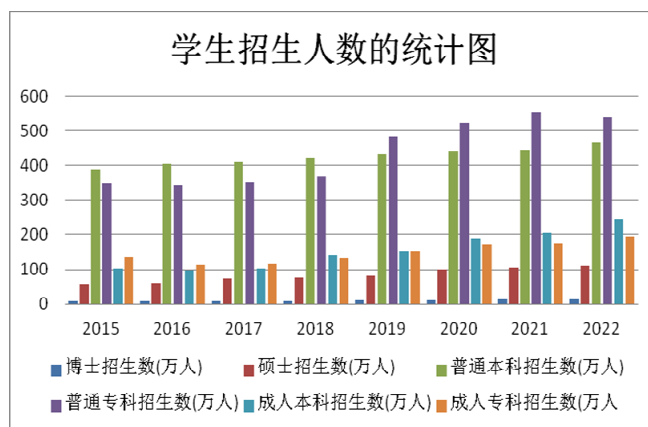
3.1.3 2015-2022年国内生产总值的统计图,如下图所示:



3.1.4 2015-2022年按产业就业人数统计图,如下图所示:



3.1.5 2015-2022年各类学生毕业的人数统计图,如下图所示:



从上面的一系列数据,我们不难看出,现阶段社会各个阶层的劳动力充足,企业招聘员工的资源非常充分,经济发展趋势较好。

3.2 招聘信息数据的选取和清洗

3.2.1 实验数据的选取。在网上截取2019年北京11,12月份400条招聘信息,数据的主要信息包括:“城市”,“公司规模”,

“地区”，“教育层级”，“招聘专业要求”，“待遇”，“工作年限”，“发布招聘时间”等重要信息。

3.2.2数据的清洗分为输入数据的清洗和输出数据的选择。

①输入数据X的清洗。教育层级的处理：我们根据不同的教育程度，将数据划分为5个等级，依次为：博士学历为5，硕士学历为4，本科学历为3，专科学历2，其它的为1。

招聘专业要求：每条招聘要求的权重设置为1，最后累计招聘条件的权重总和。

工作年限：鉴于企业都是想招聘有工作经验的老员工，所以在招聘过程中选取要求工作年限的上限作为选取的依据。

待遇：即企业给员工的待遇，按项数计算，每项为1，最后累加。

发布招聘时间：招聘时间为11月份和12月份，我们选取的原则是越新发布的招聘信息，它的权重就越大，故11月份的权重取值为1，12月份的权重为2。

地区：我们根据2019年北京各区经济的公共预测来确定招聘公司所在区的权重，处理过程如下表所示：

	2019年公共预算收入排名	预算/100
朝阳区	536.4	5.3
海淀区	446.5	4.4
北京西城区	431.5	4.3
北京东城区	189.7	1.9
北京顺义区	165.7	1.7
北京丰台区	127.7	1.3
北京昌平区	105.6	1
北京大兴区	102.5	1
北京通州区	88.73	0.9
北京房山区	70.1	0.7
石景山区	63.4	0.6
其它区域	0	0

公司的规模：这部分的权重按照规模大小来确定，如下表所示：

公司规模	取值
少于15人	0
15-50人	1
50-150人	2
150-500人	3
500-2000人	4
2000人以上	5

②输出数据Y的处理过程。月薪：我们按照给出的招聘月薪的最低值选取，选取万元/每月。

3.3算法模型的实现过程

模型首先利用最优树深度下的分类树的算法确定招聘信息表中各元素的重要程度，然后再根据元素的重要程度确定权重，创建多元回归模型。

3.3.1最优深度的确定。针对招聘信息的数据，利用机器学习中的分类树算法，首先确定最优树的深度和误差，随后利用回归树确定合理的树的深度，最后确定影响招聘的各个元素的权重。树的深度如下图所示，从下图可以看出：树的深度为4时，较为合适。

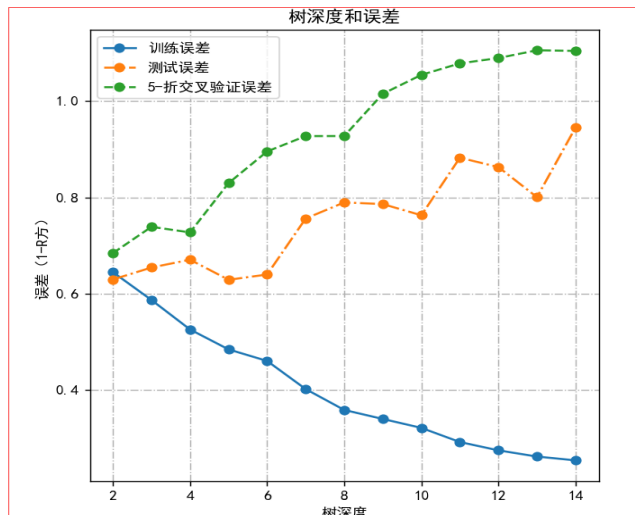


图 树的深度和误差

3.3.2利用回归树的数据模型，确定数据表中的参数的重要程度，具体结果如下图所示：

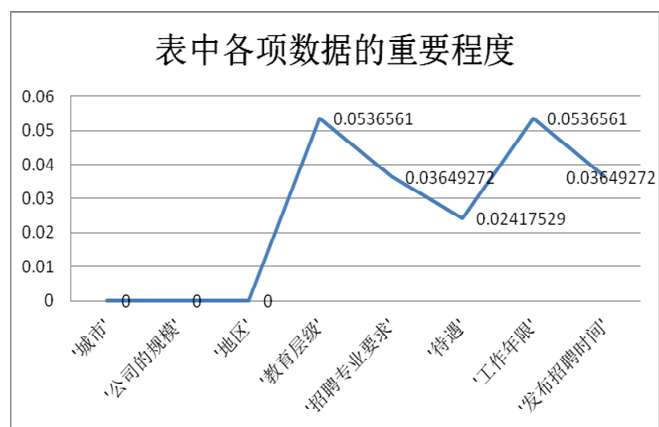


图 数据重要程度

3.3.3多元回归模型。

①回归预测模型的简介我们常用到的预测模型为：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n + \epsilon \quad \text{公式1}$$

这里我们根据上面的重要程度依次选取 β_n 分别为： β_1 教育层次、 β_3 招聘专业要求、 β_5 待遇、 β_2 工作年限、 β_4 发布招聘时间； β_0 为模型的截距项。

②损失函数 ϵ

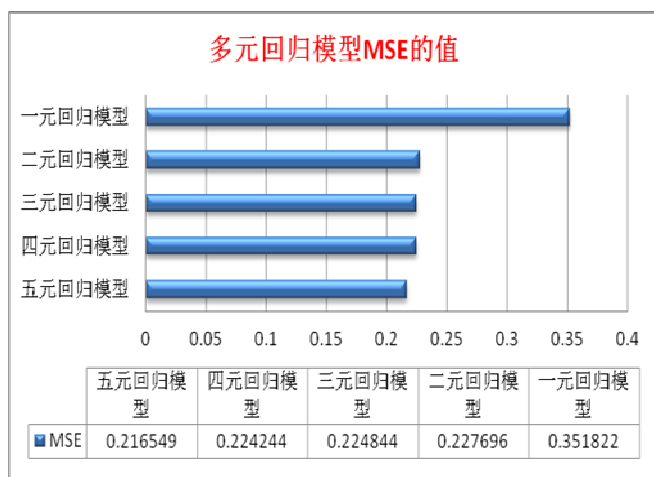
输出函数和预测函数平方误差

$$\epsilon = (y_i - \hat{y}_i)^2 \quad \text{公式2}$$

③均方差 (Mean Square Error) MSE。MSE是误差综合的评价值,也是平方损失函数的期望, MSE的值越大表示误差越大,反之则越小。

$$MSE = \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 \quad \text{公式3}$$

由上图,我们按照重要程度由大到小,选取教育层次、招聘专业要求,待遇、工作年限、发布招聘时间为输入多元回归系数,建立五元回归模型,利用五元回归模型的回归系数,确定最优的MSE,由算法可以计算出:五元回归模型的截距项:0.362779;五元回归模型的回归系数:[0.134963630.15123011-0.092878210.16336931-0.0423924];五元回归模型的MSE为:0.216549为了验证五元回归模型是最优的模型,我们根据元素的重要程度,分别计算一元回归,二元回归,三元回归,四元回归MSE的值,对比结果如下图所示:



4 总结

在信息爆炸的时代,在大数据和新技术高速发展的冲击下,人力资源管理系统和机器学习相结合成为未来发展的一种必然趋势,但目前人力资源管理还停留在传统阶段,同机器学习具体结合的应用领域不多,而且机器学习的优越性和应用价值并

未明显的体现出来。本文因此提出了一种最优深度分类树的多元回归模型算法,它能够有效地提高人力资源管理在数据分析方面的效率,明显的减少招聘人员的工作量,能为企业创造出更好的价值。

【参考文献】

[1]周志华.机器学习[M].清华大学出版社,2016.
 [2]张雪.基于文本挖掘的人力资源岗位胜任力模型构建及其应用研究[D].武汉:湖北大学,2021.
 [3]尹源.基于机器学习的企业互联网招聘中简历筛选研究[D].南京:南京邮电大学,2020.
 [4]李清黎,徐慧娟.人力资源管理系统的现状及不足[J].当代经济,2009,(06):66-67.
 [5]Ranjan J, Goyal D P,Ahson S I.Data Mining Techniques for Better Decisions in Human Resource Management Systems. International Journal of BusinessInformation Systems,2008,5: 464-481.
 [6]高超.数据挖掘技术对企业人力资源管理的影响分析[J].电子制作,2013,(17):288.
 [7]潘现玮.数据挖掘在人力资源信息分析中的应用[J].中外企业家,2016,(31):137-138.
 [8]张金艳.数据挖掘在人力资源离职管理中的应用[D].首都经济贸易大学,2016.
 [9]赖华强,王三银,仲崇高.人力资源管理领域的数据挖掘应用展望——以基于灰色关联模型的离职管理实证分析为例[J].江苏商论,2018,(08):42-47.
 [10]刘宇.基于遗传算法的模糊聚类研究及其应用[D].河南大学,2008.
 [11]姜飞.机器学习在人岗匹配度测算模型中的应用[D].湖南工业大学,2012.
 [12]Huang M,Tsou Y,Lee S.Integrating Fuzzy Data Mining and Fuzzy Artificial Neural Networks for Discovering Implicit Knowledge.Knowl-Based Syst,2006,6:396-403.