

基于逻辑回归分析的性格测试模型

韩文凤 胡正华

(南京航空航天大学经济与管理学院, 江苏 南京 211100)

摘要 文章针对九型人格测试问题,在统计学习方法和计算机技术的基础上,从基于逻辑回归分析对九型人格测试进行模型构建的角度开展研究。研究结果表明,文章提出的基于逻辑回归分析的九型人格测试方法能够有效地提高九型人格分类的准确性,同时能够提高测试的效率。

关键词 逻辑回归分析;九型人格;统计学习方法

0 引言

九型人格,也被称为性格型态学。九型人格表现出了每个人内在的最深入、最核心的价值观和最集中的注意力焦点,可以帮助企业员工明白自己的个性,了解自身性格优势和局限,为企业的人力资源管理提供有效、可靠的科学依据。目前,九型人格的测试划分方法主要有2种:量表测试和访谈测试。九型人格是动态变化的,且对应的是潜意识领域,普通的量表测试设计难度较高,信度和效度相对较低,且填写的效率很低,不利于保存和推广。访谈测试要求访谈者一定要经过专业培训,并且最好有专家参与,在企业中推广难度较大。本文研究目的是通过研究性格测试的理论和方法,运用逻辑回归分析和特征选择等模型方法构建九型人格测试模型。

1 实验数据收集与分析

实验数据的准确性和有效性直接决定了实验的结果和系统的测试效果,为了获取有效的实验数据,本文以唐理查德和里索的144道九型人格测试题为问卷,实验对象为之前系统学习过九型人格、明确自身性格特征的全国各地不同年龄、不同职业的九型人格导师、学员和爱好者。经过汇总,共收回300份问卷,每道题选择第1个选项记为-1,选择第2个选项记为1。

2 性格测试系统模型构建

逻辑回归模型是以某种结果发生的概率为因变量,影响该结果发生的因素为自变量建立回归模型。九型人格模型的逻辑回归方程如下:

$$y_i = a_0 + a_1 X_{i1} + a_2 X_{i2} + \cdots + a_{144} X_{i144} \quad (1)$$

其中 y_i —输出值,本文指类别概率,即九个型号的概率 $j=$

1, 2, ..., 9;

a_0 —回归常数;

a_j —第 j 个变量的回归系数($j=1, 2, \dots, 144$);

X_{ij} —输入值,每道题目选第1个选项取-1,否则取1; $X_1,$

X_2, \dots, X_{144} 是特征,由于每个题目有2个值,所以是288维特征,分别代表144道题的每道题目的输入值, y 代表9种型号类别的概率。

3 参数估计

3.1 实验步骤

3.1.1 数据预处理

为了提高数据的有效性和实验质量,首先对300份问卷进行数据预处理。

3.1.2 通过Python语言编程,实现逻辑回归分析+L2正则项的统计学习过程

从统计学习中贝叶斯学派的角度来说,L2正则项是对应参数的高斯先验。由于目前数据量太少,特征维数相对于训练样本数量来说,过于庞大,仅仅凭借样本数据很难直接训练出鲁棒的参数。所以,首先对模型的参数添加先验信息。

论文基于传统的打分方式,确定一组逻辑回归的初始参数,把这一组参数为均值的高斯分布作为逻辑回归参数的先验分布,常用的L2正则化通常以0为中心,而论文则以这一组参数为中心进行对逻辑回归添加L2正则约束。比如第1道题的2个选项,第1个选项对应4,第2个选项对应6。这2个选项对应2个特征,则设定:第1个特征的逻辑回归的初始值为(0.0, 0.0, 0.0, μ , 0.0, 0.0, 0.0, 0.0, 0.0),第2个特征的逻辑回归的初始值为(0.0, 0.0, 0.0, 0.0, 0.0, μ , 0.0, 0.0, 0.0),其中 μ 是高斯分布 $N(\mu, \sigma^2)$ 的参数。实验过程是通过调节 μ 和L2正则项参数 λ 来获得最优的逻辑回归参数。

3.1.3 代码基于python的numpy库和theano库实现

实验目的是通过将损失函数最小化的方式得到模型中的参数。使损失函数最小化的方法一般采用梯度下降法。

3.1.4 进行参数寻优

实验数据共300个点,由于数据量较少,使用15折交叉验证方法。实验将对 λ 的值反复进行15折交叉验证,最终选择

交叉验证过程中差值最小的那个取值, 然后根据所得到的 λ 值, 对所有的数据重新进行模型拟合。

3.2 实验结果

通过以上实验步骤, 实验结果为: 最终 λ 参数选择为 0.2, μ 的参数选择为 1.0。得到的分类错误率为 41.11%, 原始分类方式的错误率为 48.33%, 取得了显著的效果提升。

实验结果分析: 原来的九型人格测试中, 每道题目只做简单的加分, 相当于公式(1)中的回归系数均为 1, 在论文的逻辑回归实验中, 基于已知的测试数据, 通过实验对公式(1)中的回归系数进行了实验调整, 使得测试的准确率得到提升。在数据量有限的情况下, 实验效果已经得到了显著改善, 随着系统的推广以及测试人数的增多, 数据量将更加可观, 进行该项实验时得到的提升将更加显著, 系统的参数也将更加准确。

4 特征选择

特征选择的目的是在符合分类精度不会明显降低和分布情况不会明显改变的情况下, 找到对目标显著识别的最小特征子集。特征选择方法很简单, 但是运算量很大, 总共 144 个特征, 共需要做 $((144+1)/2) \times 144$ 次训练, 最后形成的结果是以特征集合的分类精度随已选特征的增多而不断发生变化的情况。由以上数据得到的汇总如图 1 所示。

表格数据表示的含义是: 通过特征选择发现, 第 5 题是最有区分度的, 按照区分度将原来的题目进行排序, 分别是 6, 134, 117..., 即系统会按照如上顺序出题, 当测试者做第 6 题时, 分类的精度是 0.661111, 当测试者做了第 6 题和第 134 题时, 分类的精度是 0.616667, 当做了第 6, 134, 117 题时, 分类精度是 0.577778 等, 总体趋势归于上升。

特征选择实验的结果是: 通过对原有的 144 题进行特征选择实验, 根据所得的分类精度, 对 144 题进行重新排序。排序的意义是: 由于完整做完 144 题测试时间较长, 系统管理员可视情况自行设定做题量, 如只做 70 道题、90 道题、100 道题等, 选取题目的标准则是依据上述排序来选择, 基于此形成的九型人格测试系统可以在答题的任意时间终止答题。

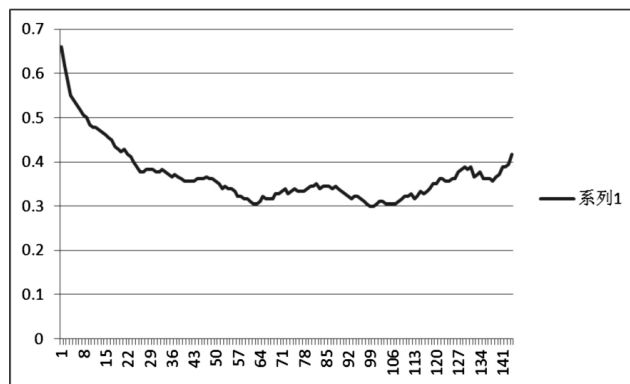


图 1 特征选择结果汇总

5 结语

本文介绍了九型人格测试模型的构建方法, 详细描述了逻辑回归分析、特征选择的方法和实验步骤, 构建了九型人格测试模型。较之前的纸质测试、访谈测试等方式, 本文为计算机化的测试过程提供了模型基础。与经典的 144 题测试相比, 由于增加了每道题目的权重设置, 同样做完 144 道题目, 本方法的准确率高于原测试题。

参考文献

- [1] 裴宇晶, 邹家峰. 九型人格与职业生涯规划[M]. 北京: 北京大学出版社, 2013.
- [2] 邹钰. 基于逻辑回归模型的垃圾短信过滤系统的研究[J]. 数字技术与应用, 2013(2).
- [3] 姚旭, 王晓丹, 张玉玺, 等. 特征选择方法综述[J]. 控制与决策, 2012(2).
- [4] 刘建伟, 付捷, 罗雄麟. L1+L2 正则化逻辑斯蒂模型分类算法[J]. 计算机工程, 2012(13).
- [5] 徐爱兵. 色彩与人格——人工神经网络在心理测验中的应用研究[D]. 苏州: 苏州大学, 2009.

(责任编辑 胡峰)

(上接第 78 页)

的黄绿色感受最亮, 光谱两端的红色和紫色觉得最暗。人眼通过有色的太阳镜看物体时, 对物体实际颜色的感受会受到影响。

3.1 灰色镜片

灰色在色谱中呈中性, 灰色镜片太阳镜在减少光强度的同时, 能使入射光仍保持原来的光谱分布, 颜色失真最小。

3.2 黄色、琥珀色和褐色镜片

很多物体会反射或散射很强的蓝光, 在表面形成一层蓝雾, 产生被称为“蓝霾”的眩光效应, 使物体界面变得模糊。黄色、琥珀色和褐色镜片对蓝光具有阻挡作用, 使物体显得明亮

而又清晰。但是, 黄色、琥珀色和褐色镜片破坏了入射光的颜色分布, 会使物体颜色产生偏差。

3.3 紫色和玫瑰色镜片

在绿色或蓝色背景下, 紫色和玫瑰色镜片能提供最好的物体对比度。因而通常用于打猎和水上运动等特殊用途太阳镜。

人们在选择太阳镜时, 不仅要考虑太阳镜的外在款式, 更应该关注其内在的质量, 并且根据不同的环境、用途正确选择太阳镜, 避免选择不当影响视力健康和人身安全。

(责任编辑 傅金睿)