

文章编号: 1004-4353(2017)04-0321-06

一种基于多层次分析和熵权的 灰色聚类评价方法

李志亮, 罗芳, 张天津
(宁德师范学院 计算机系, 福建 宁德 352100)

摘要: 在传统灰色聚类评价方法的基础上,采用多层次分析法建立评价指标体系,并利用熵权法计算聚类指标的指标权重,建立了一种基于多层次分析和熵权的灰色聚类评价方法. 实验仿真和实例验证结果表明,该评价方法具有较高的有效性和客观性,优于传统评价方法和文献[12]的评价方法.

关键词: 层次分析; 熵权法; 指标权重; 灰色聚类

中图分类号: N941.5

文献标识码: A

A gray clustering evaluation method based on multi-level analysis and entropy weight

LI Zhiliang, LUO Fang, ZHANG Tianjin
(*Department of Computer Science, Ningde Normal University, Ningde 352100, China*)

Abstract: On the basis of the traditional gray clustering evaluation method, the evaluation index system and entropy weight method are used to calculate the index weight of clustering index, and a grey clustering evaluation method based on multi level analysis and entropy weight is established. The experiment simulation and the example verification show that the grey clustering evaluation method has higher validity and objectivity, and is superior to the traditional evaluation method and the evaluation method in the literature [12].

Keywords: analytic hierarchy process; entropy weight method; index weight; gray clustering

评估过程中指标体系和指标权重对评估结果有着重要的影响,近年来一些学者采用多层次分析法建立了评估指标体系,并应用到教育、医疗、农业、经济和互联网等领域,取得了良好的评判效果^[1-4]. 自邓聚龙提出灰色系统理论^[5]以来,灰色聚类分析技术被广泛地应用于环境、军事、经济、交通和生物等多个领域. 在灰色理论的聚类分析中,学者们对指标权重和指标体系的确定进行了研究,并取得了较好的研究结果. 例如:陈启月对评测指标权重进行改进,给出了一种新的结构熵权法^[6];董一哲等采用离差最大化法计算灰色权重,给出了一种离差最大化法的灰色权重计算方法^[7];陈伟等设计了一种综合主、客观权重信息的最优组合赋权方法,有效消除了评估过程中的主观性^[8]. 在上述相关研究中,学者们或是采用层次分析法设计指标体系,或是采用熵权或优化熵权法计算指标权重,没有将这两种方法有效地融合起来. 基于此,本文针对互联网信息安全风险评估问题,采用多层次分析法建立指标体系,在聚类过程中采用熵权法计算指标权重,通过对灰色聚类的白化函数进行改进,设计了一种基于多层次分析和熵权的灰色聚类评价方法,并通过仿真实例和案例分析验证了本文方法的有效性.

收稿日期: 2017-09-12

作者简介: 李志亮(1981—),男,讲师,研究方向为不确定系统理论.

基金项目: 福建省自然科学基金资助项目(2017J01775);福建省中青年教师教育科研项目(JAT160543);宁德市指导性科技计划项目(20160103)

1 互联网信息安全风险评估的多层次分析法指标体系的建立

风险评估^[9]是互联网信息安全风险管理的重要组成部分,在互联网信息安全风险评估中需要评估的目标一般有多个,因此互联网信息安全风险评估是一个多目标决策问题.采用层次分析法可以对评估目标和规范设计统一的计量准则^[10],因此可有效地解决评估过程中的目标和规范缺乏统一计量准则的问题.层次分析法主要采用分解的思想对评估目标的评价指标系统进行分解,其基本层次有目标层、准则层和指标层,其中目标层是评价方法中评估的目标;准则层是评价方法中依据的标准和准则,往往由多个准则组成;指标层则是评价方法中各项具体的评价指标.根据互联网的信息特点,本文将系统资源、系统状态、网络流量、用户信息、威胁性和脆弱性等 6 个因素作为互联网信息安全风险的要素组成,各因素所包含的内容如图 1 所示.

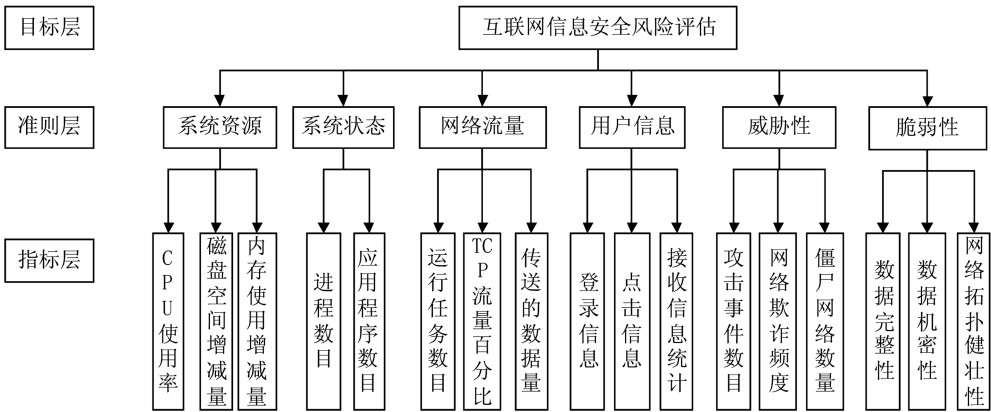


图 1 多层次分析的互联网信息安全风险评估指标体系

2 灰色熵权的计算

对评估对象进行灰色聚类时,需要计算评估对象的各个指标权重.由于熵权法能够克服多指标评价中主观确定权重的不确定性,因此本文采用熵权法计算评估指标的权重.假设对某互联网信息安全风险进行评估,评价专家有 m 个,评价指标有 n 个, i 表示参与评分的专家标号, j 表示第 j 个指标,则计算权重值的步骤如下:

1) 对数据进行预处理.对第 i 个专家对第 j 个指标的评分值 M_{ij} 进行预处理,得到处理结果用 Y_{ij} 表示,其计算公式为

$$Y_{ij} = M_{ij} / \sum_{i=1}^m M_{ij} . \tag{1}$$

2) 数据预处理后,根据熵的定义,可以得到评价指标的熵值 e_j :

$$e_j = - \frac{1}{\ln n} \sum_{i=1}^m Y_{ij} \ln Y_{ij} . \tag{2}$$

3) 根据熵值 e_j 得到第 j 个评价指标的差异系数 g_j :

$$g_j = 1 - e_j . \tag{3}$$

4) 根据差异系数 g_j 得到第 j 个评价指标的权重 ω_j :

$$\omega_j = g_j / \sum_{j=1}^n g_j . \tag{4}$$

3 基于多层次分析和熵权的灰色聚类评价方法

灰色聚类的评价过程一般分为构建指标矩阵、确定灰类数和灰类的转折点与中心点、构建白化函数

模型、确定指标权重和计算聚类系数并划分灰类. 基于多层次分析和熵权的灰色聚类评价方法的具体过程如下:

1) 在多层次分析指标体系的基础上构建指标矩阵. 设评价专家为 m 个, 评价指标为 n 个, i 表示参与评分的专家标号, j 表示评价指标编号, 对第 i 个专家对第 j 个指标的评分值用 M_{ij} 表示, 由此得到灰色聚类的评价指标矩阵 \mathbf{M} :

$$\mathbf{M} = \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1n} \\ M_{21} & M_{22} & \cdots & M_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ M_{m1} & M_{m2} & \cdots & M_{mn} \end{bmatrix} = (M_{ij})_{m \times n}. \quad (5)$$

2) 确定灰类数和灰类的转折点和中心点. 根据灰色聚类评价过程, 确定要划分的灰类数用 s 表示, 第 1 个灰类和最后 1 个灰类分别用转折点 λ_j^1 和 λ_j^s 表示, 中间第 k 个灰类的中心点用 λ_j^k 表示.

3) 构建白化函数模型. 文献[11]的研究表明, 灰类 1 和灰类 s 的三角白化权函数分别取为下限测度白化权函数和上限测度白化权函数时, 能够避免各聚类指标的取值范围向左、右延拓所带来的困扰. 本文采用改进的白化函数构建模型, 具体公式如下:

$$f_j^1(x_{ij}) = \begin{cases} 0, & x_{ij} \notin [\lambda_j^0, \lambda_j^2], \\ 1, & x_{ij} \in [\lambda_j^0, \lambda_j^1], \\ (\lambda_j^2 - x_{ij}) / (\lambda_j^2 - \lambda_j^1), & x_{ij} \in [\lambda_j^1, \lambda_j^2]; \end{cases} \quad (6)$$

$$f_j^s(x_{ij}) = \begin{cases} 0, & x_{ij} \notin [\lambda_j^{s-1}, \lambda_j^{s+1}], \\ (x_{ij} - \lambda_j^{s-1}) / (\lambda_j^s - \lambda_j^{s-1}), & x_{ij} \in [\lambda_j^{s-1}, \lambda_j^s], \\ 1, & x_{ij} \in [\lambda_j^s, \lambda_j^{s+1}]; \end{cases} \quad (7)$$

$$f_j^k(x_{ij}) = \begin{cases} 0, & x_{ij} \notin [\lambda_j^{k-1}, \lambda_j^{k+1}], \\ (x_{ij} - \lambda_j^{k-1}) / (\lambda_j^k - \lambda_j^{k-1}), & x_{ij} \in [\lambda_j^{k-1}, \lambda_j^k], \\ (\lambda_j^{k+1} - x_{ij}) / (\lambda_j^{k+1} - \lambda_j^k), & x_{ij} \in [\lambda_j^k, \lambda_j^{k+1}]. \end{cases} \quad (8)$$

其中, 为了增加聚类指标的取值范围, λ_j^s 为最后 1 个灰类转折点, λ_j^{s+1} 为 λ_j^s 的合理取值, 可以延伸为无穷大, 即大于 λ_j^s 的聚类指标值其对应的白化函数值均为 1. 在公式(8)中, $k \in [2, s-1]$.

4) 利用熵权法确定指标权重. 用 ω_j 表示第 j 个指标在评价体系中的权重值.

5) 计算聚类系数并划分灰类. 根据灰色聚类的聚类系数公式计算评价对象属于各灰类的聚类系数 σ_i^k , 计算公式为

$$\sigma_i^k = \sum_{j=1}^n f_j^k(x_{ij}) \times \omega_j. \quad (9)$$

然后根据计算得出的聚类系数 σ_i^k , 按照最大隶属度原则划分聚类对象所属的灰类.

4 仿真实验

为验证本文方法预测结果的准确性和客观性, 本文采用传统灰色聚类评价方法、文献[12]方法和本文提出的评价方法进行对比仿真实验. 实验节选 2016 年 1 月到 8 月某市 10 个电子商务网站的信息安全风险事故数据作为实验样本, 具体数据如表 1 所示. 实验聘请了 6 名专家对这 10 个电子商务网站的信息安全风险进行评分, 划分的评价结果灰类数为 4 个, 分别是很高、高、中和低. 由于灰色预测是在小样本数据的基础上进行预测的, 因此本文不考虑预测方法的时间效率和空间效率, 只考虑预测方法的准确性和客观性. 3 种方法对 10 个电子网站的信息安全风险评估结果如表 2 所示.

表 1 10 个电子商务网站的信息安全风险事故数

网站序号	1	2	3	4	5	6	7	8	9	10
事故数	9	21	29	24	22	26	15	28	27	39

表 2 3 种方法对 10 个电子商务网站信息安全风险评价的结果

测试方法	评价结果									
	1	2	3	4	5	6	7	8	9	10
传统方法	低	中	高	中	高	中	低	高	高	很高
文献[12]方法	低	中	高	中	中	中	低	高	高	很高
本文方法	低	中	高	中	中	中	低	高	中	很高

由表 2 可知,3 种方法的评价结果存在一定差异,其中主要不同的是对网站 5 和网站 9 的评价结果.网站 5 发生事故数为 22,采用传统方法的评价结果为风险等级“高”,采用文献[12]和本文评价方法的评价结果为风险等级“中”.在对网站 9 的评价结果中,本文方法的评价结果为风险等级“中”,而传统方法和文献[12]的评价方法为风险等级“高”.

为了分析上述 3 种评价方法的客观性,本文进一步对这 10 个电子商务网站的信息安全风险评估进行评分,评分采用 10 分值,分值越大表示风险值越低.评分选取两个群体:一个是电子商务网站的 10 名工作人员,一个是电子商务网站的 30 名客户,两个群体的评分结果均为去掉最高分和最低分后取平均值.两个群体的评分结果如图 2 所示.

由图 2 可知,2 个群体的评价结果均具有较高的主观性,这是由于 2 个评价群体基于不同考虑而得出的.文献[12]方法和本文方法对于电子商务网站的评价结果在工作人员和用户评价结果之间,因此说明这 2 种方法都具有较高的客观性;但本文方法比文献[12]方法的波动性更小,即评价结果受到主观性的影响更小,由此说明本文方法具有更好的客观性.

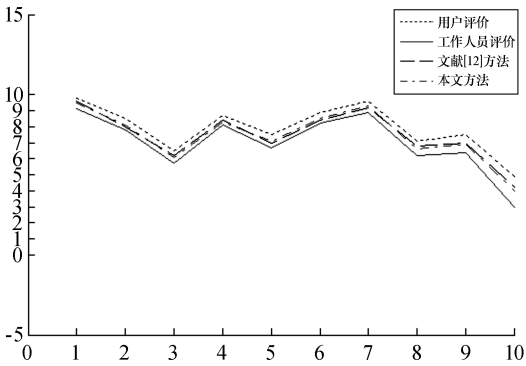


图 2 2 个群体的评价结果

5 实例分析

为了验证基于多层次分析和熵权的灰色聚类评价方法的有效性和正确性,对某电子商务网站的信息安全风险进行评价,由 5 名评价专家分别对多层次分析法指标体系中的评价指标进行评分,满分为 10 分,评分结果如表 3 所示.

表 3 专家评分结果表

专家编号	CPU 使用率	磁盘空间增减量	内存使用增减量	进程数目	应用程序数目	运行任务数量	TCP 流量百分比	传送的数据量	登录信息	点击信息	接收信息统计	攻击事件数目	网络欺诈频度	僵尸网络数量	数据完整性	数据机密性	网络拓扑健壮性
1	7.9	7.5	7.8	8.0	9.5	6.5	8.1	6.3	4.7	8.4	5.6	8.2	8.6	6.7	8.5	9.1	9.2
2	7.5	8.0	7.6	7.8	9.2	6.4	8.3	6.4	5.2	8.7	5.9	7.9	8.3	6.2	8.7	8.9	9.0
3	7.2	7.8	7.2	8.2	9.6	6.7	8.6	6.6	5.5	8.9	6.1	8.3	8.5	6.9	8.8	9.3	9.6
4	8.1	7.2	7.1	7.8	9.1	6.3	7.9	5.8	4.6	8.5	5.1	7.9	7.9	6.0	8.1	8.7	8.9
5	8.3	8.1	7.9	8.4	9.7	7.0	8.6	7.1	6.0	8.9	6.5	8.5	8.8	7.2	8.6	9.2	9.6

以表 3 数据按灰色聚类评价方法对该电子商务网站的信息安全风险进行评价. 评价过程如下:

1) 建立指标矩阵. 根据表 3 数据, 利用公式(5)建立多层次分析的指标矩阵 M :

$$M = \begin{bmatrix} 7.9 & 7.5 & 7.8 & 8.0 & 9.5 & 6.5 & 8.1 & 6.3 & 4.7 & 8.4 & 5.6 & 8.2 & 8.6 & 6.7 & 8.5 & 9.1 & 9.2 \\ 7.5 & 8.0 & 7.6 & 7.8 & 9.2 & 6.4 & 8.3 & 6.4 & 5.2 & 8.7 & 5.9 & 7.9 & 8.3 & 6.2 & 8.7 & 8.9 & 9.0 \\ 7.2 & 7.8 & 7.2 & 8.2 & 9.6 & 6.7 & 8.6 & 6.6 & 5.5 & 8.9 & 6.1 & 8.3 & 8.5 & 6.9 & 8.8 & 9.3 & 9.6 \\ 8.1 & 7.2 & 7.1 & 7.8 & 9.1 & 6.3 & 7.9 & 5.8 & 4.6 & 8.5 & 5.1 & 7.9 & 7.9 & 6.0 & 8.1 & 8.7 & 8.9 \\ 8.3 & 8.1 & 7.9 & 8.4 & 9.7 & 7.0 & 8.6 & 7.1 & 6.0 & 8.9 & 6.5 & 8.5 & 8.8 & 7.2 & 8.6 & 9.2 & 9.6 \end{bmatrix}.$$

2) 确定灰类数和灰类的转折点与中心点. 为了分析互联网信息安全, 采用 10 分制将其风险划分为 4 个等级, 划分的灰类数为 4, 如表 4 所示. 然后采用取平均值的方法确定灰类的转折点与中心点, 则“低”灰类和“很高”灰类的转折点分别是 9.5 和 2.5; “中”灰类的中心点为 8; “高”灰类的中心点为 6.

表 4 互联网信息安全风险等级划分

风险等级	很高	高	中	低
取值范围	$0 \sim \leq 5$	$5 \sim \leq 7$	$7 \sim \leq 9$	$9 \sim 10$

3) 构建白化函数模型. 根据公式(2)—(4)建立白化函数, 具体如下:

$$f_j^1(x_{ij}) = \begin{cases} 0, & x_{ij} \notin [0, 5], \\ 1, & x_{ij} \in [0, 2.5], \\ \frac{5 - x_{ij}}{5 - 2.5}, & x_{ij} \in [2.5, 5]; \end{cases} \quad f_j^2(x_{ij}) = \begin{cases} 0, & x_{ij} \notin [5, 7], \\ \frac{x_{ij} - 5}{6 - 5}, & x_{ij} \in [5, 6], \\ \frac{7 - x_{ij}}{7 - 6}, & x_{ij} \in [6, 7]; \end{cases}$$

$$f_j^k(x_{ij}) = \begin{cases} 0, & x_{ij} \notin [7, 9], \\ \frac{x_{ij} - 7}{8 - 7}, & x_{ij} \in [7, 8], \\ \frac{9 - x_{ij}}{9 - 8}, & x_{ij} \in [8, 9]; \end{cases} \quad f_j^s(x_{ij}) = \begin{cases} 0, & x_{ij} \notin [9, 10], \\ \frac{x_{ij} - 9}{9.5 - 9}, & x_{ij} \in [9, 9.5], \\ 1, & x_{ij} \in [9.5, 10]. \end{cases}$$

4) 利用熵权法确定指标权重. 按式(1)—(4) 计算评价指标体系中一级指标的权重, 结果如表 5 所示.

表 5 熵权法计算得到的一级评价指标权重值

一级指标	系统资源	系统状态	网络流量	用户信息	脆弱性	威胁性
权重值	0.175 129	0.132 715	0.162 055	0.149 894	0.176 193	0.204 013

5) 计算聚类系数并划分灰类. 根据公式(9) 计算评价对象的聚类系数 σ_i^k , 计算结果如表 6 所示. 采用最大隶属度原则划分该网站信息安全风险的灰色聚类评价结果, 由表 6 中的聚类系数数据可以看出, 该电子商务网站的信息安全风险应划分为灰类“中”, 即其风险系数一般. 这一结果与该电子商务网站的实际安全状况相符, 由此表明采用基于多层次分析和熵权的灰色聚类评价方法是可行有效的.

表 6 评价对象划分为不同风险等级的聚类系数

风险等级	很高	高	中	低
聚类系数	0.014 1	0.186 5	0.628 9	0.170 4

6 结论

采用多层次分析方法建立了互联网信息安全风险评估的指标体系, 并利用熵权法确定了指标权重, 进

而建立了一种灰色聚类白化函数模型,通过计算聚类系数来划分聚类对象所属的灰类.实验仿真和实例分析表明,本文评价方法客观、准确,有效降低了评价结果的主观性.本文在研究中只考虑了相邻区间对聚类系数的影响,并未考虑不相邻区间对聚类系数的影响,因此在后续研究过程中将考虑构建一种新的白化函数模型,以综合考虑相邻区间和不相邻区间对聚类系数的影响,进而拓展灰色聚类方法的应用范围.

参考文献:

[1] 卢皓,刘全胜,王帅帅,等.基于 AHP 的某新型坦克射击模拟训练系统成绩评定[J].四川兵工学报,2011,32(5):146-148.

[2] 陆海,王新民,赵凯瑞,等.基于灰联层次分析法的无人机自修复控制策略评价[J].航空学报,2010,31(5):1030-1037.

[3] 唐正,孙超,刘宗伟,等.基于灰色层次分析法的对抗系统效能评估[J].兵工学报,2013,34(10):1250-1257.

[4] 武兆斌,陈黎,赵春霞.基于灰色聚类和层次分析的模拟训练成绩评定[J].系统仿真学报,2012,28(2):416-424.

[5] 邓聚龙.灰理论基础[M].武汉:华中科技大学出版社,2002:1-46.

[6] 陈启月.评测指标权重确定的结构嫡权法[J].系统工程理论与实践,2010,30(7):2225-2228.

[7] 董一哲,党耀国.基于离差最大化法的灰色聚类方法[J].系统工程理论与实践,2009,29(9):141-146.

[8] 陈伟,夏建华.综合主、客观权重信息的最优组合赋权方法[J].数学的实践与认知,2007,37(1):17-21.

[9] 高阳,罗军舟.基于灰色关联决策算法的信息安全风险评估方法[J].东南大学学报,2009,39(2):225-229.

[10] Liu S F, Sheng K Q, Jeffrey Forrest. On uncertain systems and uncertain models[J]. Kybernetes: The Int J of Cybernetics, Systems and Management Sciences, 2012,41(5):548-556.

[11] 刘思峰,方志耕,杨英杰.两阶段灰色综合测度决策模型与三角白化权函数的改进[J].控制与决策,2014,29(7):1232-1238.

[12] 吉琨,王海燕,王汝传.一种基于灰色系统理论的主观信任评估方法[J].计算机技术与发展,2010,20(4):109-113.

~~~~~  
(上接第 301 页)

[3] Agarwal R P. Certain fractional  $q$ -integrals and  $q$ -derivatives[J]. Proc Cambridge Philos Soc, 1969,66:365-370.

[4] Atici F M, Elloe P W. Fractional  $q$ -calculus on a time scale[J]. J Nonlinear Math Phys, 2007,14(3):333-344.

[5] Rajkovic P M, Marinkovic S D, Stankovic M S. Fractional integrals and derivatives in  $q$ -calculus[J]. Discrete Math, 2007,1(1):311-323.

[6] 孙明哲,韩筱爽.一类分数阶  $q$ -差分边值问题的正解[J].延边大学学报(自然科学版),2013,39(4):252-255.

[7] Ahmad B, Nieto J J, Alsaedi A, et al. Existence of solutions for nonlinear fractional  $q$ -difference integral equations with two fractional orders and nonlocal four-point boundary conditions[J]. Journal of the Franklin Institute, 2014, 351:2890-2909.

[8] Agarwal R P, Ahmad B, Alsaedi A, et al. Existence theory for  $q$ -antiperiodic boundary value problems of sequential  $q$ -fractional integro differential equations[J]. Abstract and Applied Analysis, 2014,2014:1-12.

[9] Sitthiwiratham Thanin. On nonlocal fractional  $q$ -integral boundary value problems of fractional  $q$ -difference and fractional  $q$ -integrodifference equations involving different numbers of order and  $q$  [J]. Boundary Value Problems, 2016,12:1-19.

[10] Wang G, Ahmad B, Zhang L. A coupled system of nonlinear fractional differential equations with multipoint fractional boundary conditions on an unbounded domain[J]. Abstract and Applied Analysis, 2012,2012:1-11.

[11] Su Xinwei, Zhang Shuqin. Unbounded solutions to a boundary value problem of fractional order on the half-line [J]. Computers and Mathematics with Applications, 2011,61:1079-1087.

[12] Zhao Y L, Chen H B, Zhang Q M. Existence and multiplicity of positive solutions for nonhomogeneous boundary value problems with fractional  $q$ -derivatives[J]. Boundary Value Problems, 2013,2013:103.

[13] 葛琦,侯成敏.一类分数阶差分方程边值问题多重正解的存在性[J].东北石油大学学报,2012,36(4):101-110.