

基于信息熵的企业投资决策绩效评价方法

李 岚, 桑 琳, 薛 婷

(大连海事大学数学系, 辽宁大连 116026)

摘要: 企业的投资决策绩效评价是个复杂的系统, 受诸多因素影响。针对我国现行企业绩效评价体系还存在指标权重失衡的问题, 本文应用模糊聚类技术和粗糙集理论, 将大量投资项目样本数据分别去除各个属性后进行模糊聚类, 再通过对各个属性条件熵的求解, 利用信息熵知识计算各属性间的互信息, 从而确定各属性的重要性, 最后计算各属性之间的权重, 从而确定出最优的投资项目。与其它方法比较, 该方法的显著特征是无需提供问题处理数据集之外的任何先验信息, 实例仿真证明该方法是一种基于客观事实的评价方法, 企业绩效评价结果是满意的。

关键词: 应用数学; 粗糙集; 信息熵; 绩效评价; 模糊聚类; 权重

中图分类号: TP18 **文献标识码:** A **文章编号:** 1674-2850(2008)07-0533-5

Performance measurement for investment decision-making based on information entropy

LI Lan, SANG Lin, XUE Ting

(Department of Mathematics, Dalian Maritime University, Dalian, Liaoning 116026)

Abstract: Performance evaluation of enterprise is a complicated system, which is effected by many elements. This paper introduced a method of getting the weight allocation based on the information entropy from practical data, so as to solve the problem of performance evaluation of enterprise using rough set theory and fuzzy aggregation. Comparing with other methods, it has the distinguishing characteristic that it needs nothing more known information. It's proved that the performance evaluation method for investment decision-making based on information entropy of rough set was devise.

Key words: applied mathematics; rough set; information entropy; performance measurement; fuzzy aggregation; weight

0 引言

在现代企业中, 企业绩效评价在促进企业改善经营管理、提高经济效益方面具有不可替代的作用。相对于国外发达国家企业绩效评价的发展, 我国还处于起步阶段, 在我国建立市场经济的条件下对企业实施绩效评价, 具有重大的意义。

企业的投资决策绩效评价是个复杂系统, 受多因素影响^[1]。从总体上可将目前国内外常用的综合评价方法分为专家评价法、经济分析法及运筹学等; 数学方法包括数据包络法 (DEA)、模糊综合评价法、层次分析法及数理统计方法等; 现有系统的量化方法大多数都是数理统计方法, 如: 回归分析、方差分析及主成分分析等。用灰色聚类分析法可以避免上述方法要求大样本和样本有较好的分布规律、计算工作量大、出现的结果与定性分析结果不符等问题。灰色聚类分析法虽然有如上优点^[2], 但是在聚类之前的各评价指标权重设定具有很大的主观性, 我国现行企业绩效评价体系还存在一些不足, 如指标权重失衡等。

作者简介: 李岚 (1980—), 硕士研究生, 主要研究方向: 数据挖掘、智能计算

通信联系人: 桑琳, 教授, 主要研究方向: 数据挖掘、智能计算, E-mail: sanglin4@sina.com

鉴于我国企业绩效评价计分方法以功效系数法为主,综合分析判断法为辅,如果指标权重不合理,将会影响对一个企业的评价结果。在绩效评价方法确定方面,各财务指标的权重都是根据主观或经验判定事先设定好的,难免会给评价结果带来主观性,导致对某一因素过高或过低地估计,使评价结果不能完全反映公司的真实情况^[3]。定性分析指标评价的基础数据主要采用德尔菲法(专家评分法)获得,易受专家个人偏好等人为因素影响,评价结果在一定程度上带有主观随意性,从而使评价目标发生偏离。本文在文献[4]的基础上,将各个属性的权系数确定问题转化为粗糙集中属性重要性评价问题,并且通过全属性聚类 and 分别去掉各属性后聚类,计算其条件熵。由此给出各个评价指标的权重,具有一定的客观性。

1 信息熵

1965年 ZADEH 创立的模糊集理论与 1982年 PAWLAK^[5]倡导的粗糙集理论是处理不确定性问题的两种很好的方法,但两者的着眼点不同。前者研究的是同一类的不同集合中对象间的不可辨识关系,而后者研究的是不同类中对象组成的集合之间的关系^[6]。本文首先采用模糊集理论中的模糊聚类方法,在不损失任何事例信息的基础上对集合进行分类,然后采用概率粗糙集中信息熵作为对信息源不确定性的度量,从而推导出权重的分配。相关的定义如下^[7]:

定义 1 设 R 是集合 A 上的一个二元关系,如果它是自反、对称和传递的,则它是 A 上的等价关系。

定义 2 设 R 是集合 A 上的一个二元关系,与 A 中的一个元素 a 相关的所有元素的集合被称作 a 的一个等价类,记成 $[a]_R$,即 $[a]_R = \{s | (a,s) \in R\}$ 。

在粗糙集理论研究中苗夺谦^[8]、DUNTSCH 和 GEDIGA 等从信息论的角度,建立起了知识与信息熵的关系,引入信息熵和条件熵的概念。

定义 3 知识 P (属性集合) 的信息熵 $H(P)$ 定义为: $H(P) = - \sum_{i=1}^n P(X_i) \log P(X_i)$, 其中 $P(X_i) (i=1,2,\dots,n)$ 为 P 在论域 U 上的划分 $X = (X_1, X_2, \dots, X_n)$ 上 X_i 的概率。

定义 4 知识 $Q[U|IND(Q)] = \{Y_1, Y_2, \dots, Y_m\}$ 相对知识 $P[U|IND(p)] = \{X_1, X_2, \dots, X_n\}$ 的条件熵 $H(Q|P)$ 定义为 $H(Q|P) = \sum_{i=1}^n P(X_i) \sum_{j=1}^m P(Y_j|X_i) \log P(Y_j|X_i)$, 其中, $P(Y_j|X_i)$ 是条件概率, $P(Y_j|X_i) = |Y_j \cap X_i| / |X_i|, i=1,2,\dots,n, j=1,2,\dots,m$ 。

这里将二者结合起来,从而解决多因素的权重分配问题。粗糙集理论采用 SHANNON 的信息熵作为对信息源的不确定性的度量,并根据信息论定义了知识的初始熵与条件熵,从而推导出了互信息量的概念^[9~11]。互信息量反映了一个信息源从另一个信息源获取的信息量大小。本文将利用概率粗糙集模型和模糊聚类分析相结合的方法,通过求解在全部属性条件下聚类获得的信息量(初始熵),与在删除某个属性的条件下聚类获得的信息量(条件熵)做比较,从而确定全属性分类事件和删除某属性分类事件之间的互信息量,从互信息量的含义中可以看出,删除某一属性后从初始分类中所获得的信息量大小与被删除的属性所含信息量多少成反比例。因此可将互信息量的倒数作为属性重要度,并将其权值化,作为企业投资决策综合评判中参评因素的权重系数,为企业投资决策绩效评价提供数学依据。

2 基于信息熵绩效评价方法的步骤

简略的步骤如图 1 所示。

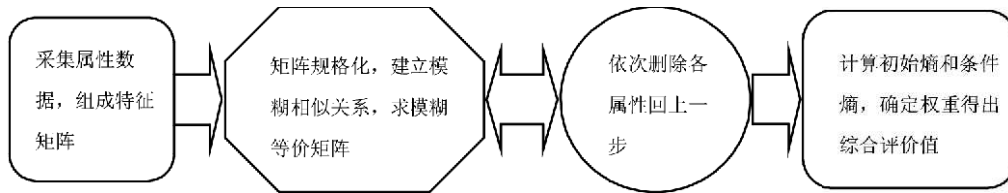


图1 绩效评价方法的步骤

Fig. 1 Steps of the way of performance measurement

2.1 数据标准化

确定需要处理的样本属性集合，采集属性数据^[4,12~13]。根据目标对样本进行初步筛选，删除掉明显不符合要求的样本，初步选取 m 个样本组成集合 $u = (u_1, u_2, \dots, u_m)$ ，每个样本有 n 个属性指标

$$X(X_1, X_2, \dots, X_n) \text{ 得到原始数矩阵为 } U = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}。$$

在实际问题中，不同的数据一般有不同的量纲，为了使不同的量纲也能进行比较，通常需要对数据做适当的变换，以消除属性值对物理量纲的影响，并将“逆向指标”正向化，使指标特征值规格在 $[0, 1]$ 范围内，规格化后的矩阵为 $R = (r_{ij})_{m \times n}$ ， $i = 1, 2, \dots, n$ ， $j = 1, 2, \dots, m$ ，规格化的方法很多^[11]。

本文采用 $r_{ij} = \begin{cases} x_{ij}/r_{\max}^i & (x_{ij} \text{ 为正向}) \\ 1 + r_{\min}^i - x_{ij}/r_{\max}^i & (x_{ij} \text{ 为逆向}) \end{cases}$ 其中 $r_{\max}^i = \max_{1 \leq j \leq m} \{x_{ij}\}$ ， $r_{\min}^i = \min_{1 \leq j \leq m} \{x_{ij}\}$ 。将规格化后的矩阵 R 转化为模糊相似矩阵， x_i 和 x_j 可采相似程度 $r_{ij} = R(x_i, x_j)$ 。确定 $r_{ij} = R(x_i, x_j)$ 的方法有若干种，本文采用 $r_{ij} = e^{-\sum_{k=1}^m |x_{ik} - x_{jk}|}$ 。

2.2 分类

利用模糊等价闭包法求模糊等价矩阵，依据模糊等价矩阵的特点确定适当阈值水平 a 后进行分类，记录分类的个数及各类中包含的元组。从全部属性中依次删除各属性后再进行步骤 2，要选择同全属性聚类相同的阈值水平，记录各类所包含的元组名称及个数，以此考察各属性对分类的影响。

2.3 求出综合评价指数

计算全属性的初始熵及分别去除各个属性后的条件熵，并确定删除各因素后的互信息量 A_j 。用 A_j 的倒数表示属性 X_j 所含信息量的相对大小， $E_j = 1/A_j$ ($j = 1, 2, \dots, n$)，将各因素所含信息量相对大小归一化处理，确定权重分配 $\omega_j = A_j / \sum_{j=1}^n A_j$ ($j = 1, 2, \dots, n$)，用规格化矩阵 B 乘以权重向量 ω 得到综合评价指数向量，值较大者为首选。

3 应用实例

企业投资决策战略绩效对公司的战略管理非常重要，在实践中参考专家的判断，根据相关行业进行指标的选取和企业实际数据的可采集情况，可以使用本文提出的方法对企业投资决策战略绩效进行评估。表 1 是某焦化公司的相关指标和数据，共选取 10 个项目用 B1~B10 表示。指标分别是投资报酬率、知识与智力资产贡献价值增长率、研发费用率、新产品产销率、相对市场占有率、职员保持率、产品生产周期效率、产品达标率、销售增长率、技术投入比率、环保投入资金率和原料的可回收率 12 个指标，用 A1~A12 表示^[2]。

为了计算方便，将逆向指标放在前三列，不改变数据信息，按照上述方法用 Matlab 编程计算。

先计算其质量规格化矩阵, 然后确定模糊等价矩阵。选取阈值水平 $\alpha=0.2944$, 全属性聚类, 分为9类: $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7\}, \{8\}, \{9, 10\}$ 。

表1 公司指标和数据
Tab.1 Index and date of the company

指标	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
	正向	正向	逆向	正向	正向	正向	逆向	正向	正向	正向	逆向	正向
B1	0.055	0.25	0.20	0.85	0.12	0.80	0.81	0.93	0.12	0.31	0.03	0.43
B2	0.085	0.30	0.25	0.9	0.08	0.87	0.75	0.91	0.14	0.35	0.04	0.51
B3	0.08	0.4	0.3	0.9	0.16	0.7	0.78	0.90	0.1	0.35	0.05	0.6
B4	0.35	0.2	0.3	0.7	0.07	0.6	0.6	0.85	0.07	0.3	0.02	0.4
B5	0.04	0.35	0.25	0.7	0.13	0.75	0.6	0.80	0.10	0.32	0.025	0.42
B6	0.08	0.56	0.35	0.85	0.16	0.85	0.56	0.80	0.12	0.36	0.03	0.5
B7	0.06	0.4	0.3	0.75	0.18	0.88	0.65	0.8	0.14	0.38	0.025	0.5
B8	0.05	0.3	0.36	0.72	0.18	0.685	0.68	0.81	0.2	0.35	0.3	0.45
B9	0.078	0.29	0.31	0.87	0.08	0.86	0.80	0.87	0.13	0.34	0.04	0.52
B10	0.075	0.30	0.32	0.79	0.12	0.73	0.79	0.85	0.11	0.36	0.04	0.51

计算得初始熵得 3.1219。

分别去掉各属性指标, 求其对应的模糊等价矩阵, 然后再进行聚类, 选取阈值水平 $\alpha=0.2944$ 。

去掉 A1 属性指标进行聚类, 分为6类: $\{1, 10\} \{2, 3, 5, 9\} \{4\} \{6\} \{7\} \{8\}$;

去掉 A2 属性指标进行聚类, 分为3类: $\{1, 2, 3, 5, 6, 7, 9, 10\} \{4\} \{8\}$;

去掉 A3 属性指标进行聚类, 分为3类: $\{1, 2, 3, 5, 9, 10\} \{4\} \{6, 7, 8\}$;

去掉 A4 属性指标进行聚类, 分为5类: $\{1, 5, 10\} \{2, 3, 9\} \{4\} \{6, 7\} \{8\}$;

去掉 A5 属性指标进行聚类, 分为3类: $\{1, 2, 3, 5, 6, 7, 9, 10\} \{4\} \{8\}$;

去掉 A6 属性指标进行聚类, 分为6类: $\{1, 2, 5, 9\} \{3\} \{4\} \{6, 7\} \{8\} \{10\}$;

去掉 A7 属性指标进行聚类, 分为3类: $\{1, 2, 3, 5, 6, 7, 9, 10\} \{4\} \{8\}$;

去掉 A8 属性指标进行聚类, 分为5类: $\{1\} \{4\} \{2, 3, 5, 7, 9, 10\} \{6\} \{8\}$;

去掉 A9 属性指标进行聚类, 分为4类: $\{1, 3, 2, 5, 9, 10\} \{4\} \{8\} \{7, 6\}$;

去掉 A10 属性指标进行聚类, 分为6类: $\{1, 3, 5, 9, 10\} \{2\} \{4\} \{6\} \{7\} \{8\}$;

去掉 A11 属性指标进行聚类, 分为6类: $\{1, 10\} \{2, 3, 5, 9\} \{4\} \{6\} \{7\} \{8\}$;

去掉 A12 属性指标进行聚类, 分为7类: $\{1\} \{2\} \{3, 5, 9, 10\} \{4\} \{6\} \{7\} \{8\}$;

依次删除各属性后的条件熵 HYX 为: $[0.4000, 0.3750, 0.4308, 0.4227, 0.3750, 0.3500, 0.3750, 0.3447, 0.3447, 0.3251, 0.4000, 0.2500]$ 。

则计算互信息量得 IXY 为: $[2.7219, 2.7469, 2.6911, 2.6992, 2.7469, 2.7719, 2.7469, 2.7772, 2.7772, 2.7968, 2.7219, 2.8719]$ 。

计算各权重分配为: $[0.0843, 0.0836, 0.0853, 0.0851, 0.0836, 0.0828, 0.0836, 0.0827, 0.0827, 0.0821, 0.0843, 0.0799]$ 。

用规格化矩阵与队员权重相乘得出绩效评价结果为: $[0.7424, 0.7674, 0.7793, 0.7181, 0.7323, 0.8175, 0.8082, 0.6866, 0.7327, 0.7243]$ 。

由以上数据可以看出, 项目绩效较好的为项目2、项目3、项目6和项目7; 绩效较差的为项目4和项目8, 与文献中运用模糊综合评价的结果相吻合。

4 结论

本文对粗糙集理论在多指标综合评价中的应用进行了讨论。分析了如何利用粗糙集进行综合评价

权重的确定,以得到更合理的权值;以及如何结合粗糙集与模糊聚类技术进行综合评价,以取得更精确的评价值。结果表明,该方法能够在不损失原始信息的情况下,用各属性所含信息量的大小来确定各属性的权重大小,从而确定各项目的综合评价指数,具有一定的客观性。可见这种评价方法是简单有效的,达到了预期的效果。

[参考文献] (References)

- [1] 王宗军, 吴文娟. 基于数据挖掘的企业投资决策绩效评价研究[J]. 科技进步与对策, 2007 (1): 56~59.
WANG Z J, WU W J. Research on investment decision-making performance evaluate of enterprise based on data-mining[J]. Science & Technology Progress and Policy, 2007(1): 56~59. (in Chinese)
- [2] 肖得云, 付智慧. 投资决策战略绩效评价[M]. 北京: 经济管理出版社, 2006.
XIAO D Y, FU Z H. Evaluation of investment decision-making performance stratagem[M]. Beijing: Economy & Management Publishing House, 2006. (in Chinese)
- [3] 黄定轩. 基于客观信息熵的多因素权重分配方法[J]. 系统工程理论方法应用, 2003, 12 (4): 321~324.
HUANG D X. Way of excessive ingredient weight distribute based on objectivity information entropy[J]. Systems Engineering Theory Methodology Applications, 2003, 12(4): 321~324. (in Chinese)
- [4] 皋军, 王建东. 一种基于模糊理论和条件熵的属性近似约简的方法[J]. 计算机工程与应用, 2004 (21): 183~212.
GAO J, WANG J D. A way of attribute approximately reduction based on fuzzy theory and condition entropy[J]. Computer Engineering and Applications, 2004(21): 183~212. (in Chinese)
- [5] PAWLAK Z. Rough sets[J]. International Journal of Computer and Information Sciences, 1982, 11: 341~356.
- [6] LIANG J, SHI Z, LI D, et al. Information entropy, rough entropy and knowledge granulation in incomplete information systems[J]. International Journal of General Systems, 2006, 35(6): 641~654.
- [7] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法[M]. 北京: 科学出版社, 2001.
ZHANG W X, WU W Z, LIANG J Y, et al. Theoretics and measures of rough set science press[M]. Beijing: Science Press, 2001. (in Chinese)
- [8] 苗夺谦, 王珏. 粗糙集理论中概念与运算的信息表示[J]. 软件学报, 1999, 10 (2): 113~116.
MAO D Q, WANG J. Conception and algorithm of informational express in rough sets[J]. Software Transaction, 1999, 10(2): 113~116. (in Chinese)
- [9] WUEST L, NICKERSON B, MUREIKA R. Information entropy of non-probabilistic processes[J]. Geographical Analysis, 2003, 35(3): 215~249.
- [10] LIANG J Y, CHIN K S, DANG C Y. A new method for measuring uncertainty and fuzziness in rough set theory[J]. International Journal of General Systems, 2002, 31(4): 331~342.
- [11] 陈水利, 李敬功, 王向功. 模糊集理论及其应用[M]. 北京: 科学出版社, 2005.
CHEN S L, LI J G, WANG X G. Theory and application of fuzzy sets science publishing company[M]. Beijing: Science Press, 2005. (in Chinese)
- [12] 王向阳. 基于近似精度和条件信息熵的粗糙集不确定性度量方法[J]. 上海交通大学学报, 2006 (7): 1130~1134.
WANG X Y. The way of uncertainty measurement based on approximately precision and conditional information entropy[J]. Journal of Shanghai Jiaotong University, 2006(7): 1130~1134. (in Chinese)
- [13] 黄洁, 刘发全. 基于粗糙集的建筑材料供应商选择方法 [J]. 物流技术, 2005 (1): 52~54.
HUANG J, LIU F Q. The way of choosing construction material supply merchant based on rough set[J]. Logistic Technology, 2005(1): 52~54. (in Chinese)