

SVM 在建设项目方案评价中的应用

朱明放

(陕西理工学院 计算机系, 陕西 汉中 723003)

摘要: 决策分析是一项生产实践中重要的复杂活动, 科学的决策需要考虑众多因素, 多指标决策分析研究这一方法。多指标决策解决的是有多个属性指标的有限决策方案的排序问题, 其关键是获取决策者的偏好信息, 多属性效用函数是解决此问题的常用方法。应用支持向量机的多属性决策方法, 对建设项目进行评价, 阐述了评价方法, 并结合实例给出实验结果。实验表明, 与其他决策方法比较, 支持向量方法作为评价分析更加有效, 它能更大区分各个评价对象, 扩大支持向量机的应用范围, 丰富多属性决策分析的内容。

关键词: 多属性决策; 支持向量机; 效用函数

中图分类号: TP18

文献标识码: A

文章编号: 1673-629X(2008)03-0243-03

Application of SVMs for Construction Scheme Evaluation

ZHU Ming-fang

(Computer Dept., Shaanxi University of Technology, Hanzhong 723003, China)

Abstract: Decision making is an important and complex activity in daily lives, a good decision needs thinking a lots of factors which affect decisions. Multiple attribute decision making is solved sorting of limited decision-making projects which have multiple attributes, it is key step to capture and represent the decision maker's preference. Multiple attribute utility function is a common method to solve the decision making projects. This paper applies the support vector machines (SVMs) method for multiple attribute decision making to evaluate the construction scheme, at same time to compare this method with other methods which solve same scheme. The study shows based on support vector machine is a good and efficient method for multiple attribute decision, as it can distinguish all candidates, thus enlarge application of SVMs, enrich contents of multiple attribute decision making.

Key words: multiple attribute decision making; support vector machines; utility function

0 引言

近年来, 多指标决策问题的研究已取得了很大的进展, 并成为运筹学、决策学和系统科学等学科中十分活跃的研究课题。多指标决策 (Multiple Attribute Decision Making, MADM), 又称多属性决策, 其主要解决的是有多个属性指标的有限决策方案的排序问题^[1]。目前已有许多解决多属性决策的排序法, 如理想点法、加权平方和法、主成分分析法、模糊综合评价、灰色关联度法等, 这些评价方法都需要在评价前确定各指标的权重, 也就是决策者的偏好。权重是指标本身物理属性的客观反映, 其大小反映了综合评判中各参评因素的相对重要程度, 取值好坏将直接影响到评价结果的好坏, 是解决多属性决策问题的关键过程。

通常用多属性效用函数 (Multiple Attribute Utility Function, MAUF) 来衡量决策者对备选方案的满意程度, 并根据效用函数值对方案集进行排序并选出最满意方案。由于多属性效用函数结构复杂, 在一般情况下较难设定它们的值。在实际应用中, 人们常简单地将多属性效用函数分解为单个属性效用函数的加性形式、乘法形式或拟加性形式。

基于支持向量机 (Support Vector Machines, SVMs) 的非线性系统建模, 不需要事先对函数关系进行任何假设。利用这个特性, 根据每个属性的效用函数值, 生成描述多属性决策问题的样本, 来训练支持向量回归机, 得到决策者的偏好结构信息, 用训练好的回归机对决策方案估计效用函数值, 实现对决策方案的排序和选择。

1 TOPSIS 多属性决策方法

理想点法也称作 TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) 法, 直译为逼

收稿日期: 2007-06-17

基金项目: 陕西理工学院科研基金资助项目 (SLG0535)

作者简介: 朱明放 (1970-), 男, 陕西咸阳人, 副教授, 博士, 研究方向为数据库与数据挖掘。

近于理想解的排序方法,这种方法通过构造多指标问题的理想和负理想解,即各指标的最优解和最劣解,并以靠近理想解和远离负理想解的程度,通过计算每个方案到理想方案的相对贴近度来对方案进行排序,从而选择出最优方案。

设决策问题有 m 个可行方案 $a_i (i = 1, 2, \dots, m)$, 有 n 个指标 $f_j (j = 1, 2, \dots, n)$, 并设各指标为效益型, 即指标属性值都需要正向化。若有其他类型指标, 可以转化为效益型。这里只考虑效益型指标和成本型指标, 按极差变换法进行标准化。具体地, 对效益型指标采用公式(1), 成本型指标用公式(2), 对指标进行规范化。

$$y_{ij} = \frac{x_{ij} - \min x_j}{\max x_j - \min x_j} \quad (1)$$

$$y_{ij} = \frac{\max x_j - x_{ij}}{\max x_j - \min x_j} \quad (2)$$

TOPSIS 法的数学模型^[2] 如下:

设决策矩阵为 $X = (x_{ij})_{m \times n}$, 指标权重向量为 $W = (w_1, w_2, \dots, w_n)^T$ 。

TOPSIS 法的基本步骤是:

(1) 对决策矩阵作标准化处理, 得到标准化矩阵:

$$Y = (y_{ij})_{m \times n}$$

(2) 计算加权标准化矩阵:

$$V = (v_{ij})_{m \times n} = (w_j \times y_{ij})_{m \times n}$$

(3) 确定理想和负理想解:

$$\text{正理想解: } V^+ = \{ \max_{1 \leq i \leq m} v_{ij} \mid j = 1, 2, \dots, n \} = \{ v_1^+, v_2^+, \dots, v_n^+ \}$$

$$\text{负理想解: } V^- = \{ \min_{1 \leq i \leq m} v_{ij} \mid j = 1, 2, \dots, n \} = \{ v_1^-, v_2^-, \dots, v_n^- \}$$

(4) 计算各方案到正理想和负理想解的距离:

对于 $i = 1, 2, \dots, m$, 方案 a_i 对正理想解的距离:

$$S_i^+ = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^+)^2}, \text{ 对负理想解的距离是 } S_i^- = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^-)^2}$$

(5) 计算各方案的相对贴近度:

$$C_i^* = (1 + \frac{S_i^+}{S_i^-})^{-1} = \frac{S_i^-}{S_i^- + S_i^+}$$

(6) 按相对贴近度的大小对各方案进行排序, 相对贴近度大者为优, 相对贴近度小者为劣。

可见, TOPSIS 法关键就是前 2 步: 标准化和权重确定, 然后通过计算, 对决策方案进行排序。在基于支持向量机决策的方法中, 根据各个属性的正负理想点, 构造决策样本, 通过学习获取决策问题的结构信息, 然后将决策方案输入到学习来的支持向量机, 得到各个方案的效用函数值, 完成决策任务。

可见, 基于支持向量机的决策方法不同于其他决策方法, 它对描述决策者的偏好信息是通过样本学习得到的, 不是通过主观赋给的, 体现了数据挖掘——“让数据自己说”的精神, 因而更客观更有效。

2 多属性决策的支持向量机方法

文献[3,4]研究多属性决策的支持向量机模型和方法。此节将其更清晰地总结为实现步骤, 以便编程和实施算法。

算法: 多属性决策的支持向量机

输入: 决策问题的原矩阵

输出: 决策方案的排列顺序

步骤:

(1) 输入决策方案矩阵 X , 并按第 1 节方法对 X 预处理, 得到矩阵 Y 。

(2) 求正负理想点:

$$\text{正理想点: } V^+ = \{ \max_{1 \leq i \leq m} y_{ij} \mid j = 1, 2, \dots, n \} = \{ y_1^+, y_2^+, \dots, y_n^+ \}$$

$$\text{负理想点: } V^- = \{ \min_{1 \leq i \leq m} y_{ij} \mid j = 1, 2, \dots, n \} = \{ y_1^-, y_2^-, \dots, y_n^- \}$$

(3) 确定每个属性正负理想点的效用函数: 令 $y_j^+ = 0.95, y_j^- = 0.05, j = 1, 2, \dots, m$ 。

(4) 生成能反映决策对象整体特征的样本: 随机样本, 方案样本和正负理想点样本。

(5) 选择部分随机样本作为训练样本集, 其余的样本作为测试集(10%左右)。选择适当核函数和反复调整核参数(可以采用参数寻优方法获取), 建立基于支持向量机的回归模型, 直到获得满意的结果。

(6) 得到训练好的支持向量回归模型, 将决策方案代入, 求得方案的效用函数值。

(7) 对方案的效用函数值排序, 选择最优方案。

与 TOPSIS 法比较, 这里的效用函数是通过学习得到的, 避免了人为赋权值的主观性, 再有这种方法不需要更多的专业领域知识, 所以有更好的实用性。

3 决策实例

3.1 数据资料和决策结果

为了便于比较文献[5]的实例, 在某建筑工程的评标过程中, 经对申请的投标商的资格审查, 核准有 9 家投标商, 分别用 A_1, A_2, \dots, A_9 表示。表 1 表示各投标商相关指标。

表 1 中, 各属性特点是: 属性施工方案, 技术力量, 管理水平, 社会信誉为效益型指标, 其余为成本型指标。为了应用文中方法, 首先按照属性特点对其正向

化和规范化。

表 1 9 家投标单位的各项指标

投标单位	报价(万元)	工期(月)	钢材用量(t)	木材用量(m ³)	水泥用量(t)	施工方案	技术力量管理水平	社会信誉
A1	2122	34	2698	2148	8722	0.7898	0.8396	0.7770
A2	2031	32	2804	1938	8044	0.7692	0.7698	0.8110
A3	2250	29	2468	2021	8704	0.6357	0.8539	0.7558
A4	2180	31	2514	2342	8541	0.6798	0.9001	0.8802
A5	2391	30	2701	2101	8604	0.5405	0.7605	0.8396
A6	2010	33	2840	1986	8286	0.5801	0.8000	0.7800
A7	2310	29	2907	2085	8800	0.5168	0.7102	0.9001
A8	2101	32	2745	2169	8714	0.7598	0.8503	0.8203
A9	2081	30	2690	2207	8201	0.8598	0.8896	0.9302

用 Matlab6.5 编程,核函数选为高斯径向核函数,参数取为 $C = 2000, \epsilon = 0.005, \sigma = 5$, 损失函数取为 ϵ -不敏感损失函数。在训练集上得到支持向量个数为 71, 占 51.4%, 随机选取 10% 样本作为测试集, 得到平均绝对误差 $MAE = 0.0041$, 均方误差 $MSE = 1.9490e-005$, 罗列一次测试的 9 个样本上的期望及预测情况见表 2。

将规范化的决策矩阵 R 输入该支持向量机, 得到各方案的效用函数值, 列表见表 3, 按效用函数值大小排序-得到决策结果为 $A9 > A2 > A4 > A3 > A8 > A6 > A1 > A5 > A7$ 。决策人根据这个结果从中再选择几个单位, 或增加一些指标以供进一步决策使用。

3.2 与其他方法的比对

文献[5]使用模糊物元模型对该问题进行了研究,

表 2 随机选取的 10% 样本作为测试集的测试情况

样本点	1	2	3	4	5	6	7	8	9
期望值	0.4700	0.3700	0.7600	0.6100	0.2400	0.3400	0.5400	0.4400	0.2300
预测值	0.4714	0.3759	0.7556	0.6043	0.2443	0.3465	0.5375	0.4430	0.2336
绝对误差	-0.0014	-0.0059	0.0044	0.0057	-0.0043	-0.0065	0.0025	-0.0030	-0.0036
相对误差	-0.00298	-0.01595	0.005789	0.009344	-0.01792	-0.01912	0.00463	-0.00682	-0.01565
MAE	0.0041								
MSE	1.9490e-005								

参考文献:

[1] 刘树林, 邱菀华. 多属性决策基础理论研究[J]. 系统工程理论与实践, 1998, 18(1): 39-43.
[2] 陈伟. 关于 TOPSIS 法应用中的逆序问题及消除的方法[J]. 运筹与管理, 2005, 10: 50-54.
[3] 王强, 沈永平, 陈英武. 多属性决策支持向量机模型与算

法[J]. 控制与决策, 2006, 21(12): 1338-1342.
[4] 王强, 沈永平, 陈英武. 多属性决策的支持向量机方法[J]. 系统工程理论与实践, 2006(6): 54-58.
[5] 隋海荣, 房栓社, 马福雷. 建设项目方案评价的模糊物元模型及其应用[J]. 基建优化, 2006(4): 54-56.
[6] 庞南生, 张洪青. 一种基于 Spearman 秩相关系数与熵的多属性决策方法[J]. 运筹与管理, 2003(5): 52-56.

文献[6]对该问题作了 Spearman 秩相关系数与熵方法的研究, 2 种方法都采用了熵确定权系数, 得到了与文中相似的结果。文献[4]的决策结果为: $A9 > A1 > A8 > A4 > A7 > A2 > A6 > A3 > A5$, 文献[3]的方案决策顺序为: $A9 > A8 > A2 > A1 > A4 > A3 > A6 > A5 > A7$ 。结果大趋势基本相同。

表 3 9 个方案的决策结果

方案	A1	A2	A3	A4	A5	A6	A7	A8	A9
效用函数值	0.3885	0.6496	0.5556	0.5908	0.3175	0.4415	0.2730	0.4764	0.8286
排序	7	2	4	3	8	6	9	5	1

还可以发现, 排序第一的 A9 方案和第二的 A2 方案的效用值分别为 0.8286 和 0.6496, 文献[3]中排序第一的 A9 方案和第二的 A8 方案的效用值分别为 0.7894 和 0.5905, 文献[4]中排序第一的 A9 方案和第二的 A1 方案的效用值分别为 0.720 和 0.714。可见, 文中的方法在选优时更具有区分能力。

4 结 论

SVM 已取得了广泛应用。文中基于多属性决策的 SVM 方法对建设项目方案进行决策分析, 利用 SVM 来估计多属性效用函数, 表达决策者的偏好结构信息。SVM 来自机器学习领域, 体现“让数据自己说”的精神, 不用考虑决策属性间的复杂关系, 通过理想点法从决策问题本身构造学习样本, 通过对样本的自学习, 实现决策的智能化。进一步扩大了支持向量机的应用范围, 也为多属性决策问题引入新的方法。