

供应商评价的支持向量机模型及应用研究*

曾绍华^{1,2}, 魏 延^{1,3}

(1. 重庆大学 自动化学院, 重庆 400044 ; 2. 重庆师范大学 管理学院 ;
3. 重庆师范大学 数学与计算机学院, 重庆 400047)

摘 要 对供应商的评价是企业供应中的首要问题。本文在建立供应关系数据仓库的基础上,挖掘和优化供应商评价指标体系。应用“温度计-洋葱头”算法建立供应商评价的隶属函数,以定义供应商评价的目标变量——供应商评价指数,和建立挖掘供应商评价的支持向量机模型。最后介绍了一个实例。

关键词 供应商评价;支持向量机;数据挖掘

中图分类号: TP182

文献标识码: A

文章编号: 1672-669X(2007)01-0029-05

Research on SVM Model of Supplier Evaluation & Application

ZENG Shao-hua^{1,2}, WEI Yan^{1,3}

(1. College of Automation, Chongqing University, Chongqing 400044, China ;
2. College of Management, Chongqing Normal University ;
3. College of Mathematics and Computer Science, Chongqing Normal University, Chongqing 400047, China)

Abstract One important issue in evaluating a business supply system is the evaluation of the suppliers involved. This paper addresses the mining and optimization of the supplier evaluation system based on the supplier data warehouse. The membership function for evaluating suppliers is established by using the thermometer-onion algorithm. On this basis, novel supplier evaluation index is introduced, and a support vector machine based on supplier evaluation system is established. The proposed model is justified by a practical application.

Key words supplier evaluation ; SVM (Support Vector Machine) ; data mining

过去,企业只注重生产技术和市场营销的研究。近年来,部分学者对成本进行分析,发现有相当部分企业从供应商采购的零件或原材料占了产品成本的50%以上,有的高达80%。因而,大多数学者认为供应、生产和销售成为推动企业前进的“三驾马车”,企业开始注重对供应的研究。

在供应“马车”中,对供应商评价是首要的问题。前人就如何评价供应商做了大量研究工作。对供应商的技术、质量及认证、价格、供货能力、服务、合作时间、地理位置、财务状况和发展战略等进行研究,建立了包括这些因素的评价指标^[1~5]。有的评价指标体系繁琐冗长,还有部分指标涉及供应商的商业机密,无法获取,不能实际参与评价。一些学者总希望这包罗万象的指标体系能全面评价供应商。

评价方法多采用数据包络、目标决策、相关分析、因子分析和回归分析等统计分析方法^[2,5,6]。也未对评价指标体系优化,评价结果泛化性不强,评价实时性差。

本文介绍了建立供应关系数据仓库,倡导从某个方面去评价供应商。从供应关系数据仓库中挖掘和优化供应商评价指标,并建立供应商评价的支持向量机模型,从而提高供应商评价结果泛化性和评价实时性。最后作为应用实例,介绍了重庆某企业与供应商经济合作和效能的评价。

1 供应关系数据仓库及评价目标变量

1.1 供应关系数据仓库的建立

供应商评价需要供应商的相关信息。企业收集

* 收稿日期 2006-03-10 修回日期 2006-09-25

资助项目:重庆市教委科学技术研究项目(No. KJ050809)

作者简介:曾绍华(1969-)男,重庆人,博士研究生,研究方向为智能算法与知识发现(KDD)。

供应商信息容易犯两类错误:一类是忽视供应商供货的相关信息。一些企业在接受供应商供货时,只在会计账目中记录供货数量和金额;供应部门也只记录供应商供货的产品和联系电话等;而产品的批次、合同供货日期、交货日期、运输、质量、价格、仓储费用和售后服务等对供应商评价有价值的信息往往被忽略掉,并且对供应商的信息未信息化。另一类是片面强调虚信息的获取。一些企业总希望获取供应商的所有信息,包括供应商的财务状况、发展战略和技术研发水平等。这些信息是企业的商业机密,供应商不会提供,即使企业收集到这类信息多数也是虚假信息,简称虚信息。

企业应立足依靠自身收集供应商信息。供应(包括运输和仓储)、质检、财务和售后服务等部门协同,建立供应商评价的基础——供应关系数据仓库。

供应关系数据仓库结构及部门责任如图1所示。



图1 供应关系数据仓库结构及部门责任

1.2 供应商评价的目标变量

供应商评价就是从供应关系数据仓库中挖掘出供应商评价模型。一般情况下,建立有监督学习模型,需要定义目标变量。

在数据挖掘中,常常需要定义诸如供应商的优劣、客户忠诚度和客户满意度等目标变量。首先可以联想到人们是如何定义评判冷热这个物理量的。人们确定冷和热是相对的,如果定义一个指标并能利用该指标评判A比B冷或热,那么就可以利用该指标来度量冷热了。温度计就是这样制成的。同理,也像定义冷热物理量那样定义数据挖掘中的目标变量。这也是下面算法名称中有“温度计”一词的原因。

但是,在数据挖掘中,影响供应商优劣、客户忠诚度和客户满意度等目标变量的因素很多,它是一

个n维向量 $X=(x_1, x_2, \dots, x_n)$ 。将 $X \in \mathbf{R}^n$ 映射到目标变量 $Y \in \mathbf{R}$,就实现了目标变量Y的定义,即 $X \mapsto Y (\mathbf{R}^n \mapsto \mathbf{R})$ 。

1.3 “温度计-洋葱头”算法

分析影响目标变量 $Y \in \mathbf{R}$ 的因素 $X=(x_1, x_2, \dots, x_n) \in \mathbf{R}^n$ 在其它条件不变的情况下,存在 x_{ij} (表示第i个样本的第j个属性)越大, Y_i 越大;存在 x_{ik} (表示第i个样本的第k个属性)越大, Y_i 越小。实际上 x_{ij} 和 x_{ik} 是一类。做如下的变换

$$X \mapsto X' (\mathbf{R}^n \mapsto \mathbf{R}^n)$$

$$\text{即} \begin{cases} x'_{ij} = x_{ij} \\ x'_{ik} = c_k - x_{ik} \end{cases} \quad (c_k \text{ 为常数}) \quad (1)$$

在 X' 中 x'_{ij} 和 x'_{ik} 完全一致。

通常,在应用的取值范围内,在其它条件不变的情况下, x'_{ij} 越大,目标变量 Y_i 越大。在下面的研究中也作这样的假设。

在实际应用中,进一步分析 x'_{ij} ,在其它条件不变的情况下,当 $x'_{ij} \leq c_{1j}$ 或 $x'_{ij} \geq c_{2j}$ (c_{1j}, c_{2j} 为常数)的时候, x'_{ij} 的变化几乎不会引起 Y_i 的变化或 Y_i 变化很小;当 $c_{2j} > x'_{ij} > c_{1j}$ 时, x'_{ij} 的变化对 Y_i 有显著影响。对 x'_{ij} 做进一步变换

$$x'_{ij} = \begin{cases} c_{1j}, & x'_{ij} \leq c_{1j} \\ x'_{ij}, & c_{1j} < x'_{ij} < c_{2j} \\ c_{2j}, & x'_{ij} \geq c_{2j} \end{cases} \quad (2)$$

现在对 x'_{ij} 做归一化处理,消除其量纲的影响,即将 x'_{ij} 映射到半径为1的超球的I卦限中,并引入模糊数学的隶属函数思想,定义 $Y = \|X'\|$ 。即Y将I卦限中的超球体化分为洋葱瓣,这也是算法名称中“洋葱头”的来历。

归纳“温度计-洋葱头”算法如下。

算法1

Step1 用公式(1)将 $X \mapsto X' (\mathbf{R}^n \mapsto \mathbf{R}^n)$;

Step2 用公式(2)变换 x'_{ij} ;

Step3 $x'_{ij} = \frac{x'_{ij} - c_{1j}}{c_{2j} - c_{1j}};$ (3)

Step4 $Y_i = \|X'_i\|。$ (4)

2 供应商评价指标体系

2.1 供应商评价指标体系的确定

从供应关系数据仓库中,能够钻取包括供应商资质认证、合同、供应、退换货、质量、货款支付、供应商的经济效能和售后服务等8大类影响供应商评价的信息,几乎包含了文献[1~5]所列出的全部类

别。

这些指标不一定能满足评价要求, 还可以衍生许多新的指标。但是, 这众多的信息就能全面地评价供应商吗? 这些信息中就没有冗余吗? 许多学者, 都试图去寻找一个全面的、没有冗余的供应商评价指标体系^[1-5]。实际上, 这是非常困难的。其实企业只对供应商的某方面感兴趣, 也只需对供应商的某方面进行评价, 没有必要去苦苦寻求那根本不存在的全面的、没有冗余的指标体系。

确定供应商评价指标体系的主要原则如下。

- (1) 明确供应商评价的侧重点;
- (2) 立足能收集的相关信息构造评价指标体系;
- (3) 对评价指标体系进行优化。

2.2 供应商评价指标体系的优化

假定企业需要对供应商的某一方面 $Y \in \mathbf{R}$ 进行评价, 从供应关系数据仓库钻取和衍生相关的 n 个指标构成 n 维向量 $X = (x_1, x_2, \dots, x_n) \in \mathbf{R}^n$, 组成新的矩阵 $A_i = (x_1, x_2, \dots, x_n, y) \in \mathbf{R}^{n+1}$ $i = 1, \dots, l$ 。

$$\text{计算}^{[7]} \text{cov}A = (A - \bar{A})(A - \bar{A})^T \quad (5)$$

相关矩阵 R 其中

$$r_{ij} = \frac{\text{cov}A_{ij}}{\sqrt{\text{cov}A_{ii}}\sqrt{\text{cov}A_{jj}}} \quad ij \in \{1, 2, \dots, n, y\} \quad (6)$$

在 R 矩阵 $r_{ij} = (r_{1y}, r_{2y}, \dots, r_{ny})$ 中, 选取 k 个较大 r 值对应的指标构成优化的评价指标体系 $X_{opt} = (x_1, x_2, \dots, x_k) \in \mathbf{R}^k$, 从而剔除与目标变量相关性差的指标, 降低评价指标体系的维数和计算支持向量机模型的复杂度。

3 供应商评价的支持向量机模型

3.1 支持向量机模型

支持向量机是有监督的学习模型^[8], 其模型为

$$Y(X) = w^T \varphi(X) + b \quad (7)$$

其中 $Y \in \mathbf{R}$ 是目标变量, $X = (x_1, x_2, \dots, x_n) \in \mathbf{R}^n$ 是影响因素。

选取结构化风险泛函为

$$J(w, \rho) = \frac{1}{2} w w^T + \frac{1}{2} \gamma \sum_{i=1}^k e_i^2 \quad (i = 1, 2, \dots, k) \quad (8)$$

约束条件为

$$y(x_i) = w^T \varphi(x_i) + b + e_i \quad (i = 1, 2, \dots, k) \quad (9)$$

建立目标规划

$$\min J(w, \rho) = \frac{1}{2} w w^T + \frac{1}{2} \gamma \sum_{i=1}^k e_i^2 \quad (10)$$

$$\text{st. } y(x_i) = w^T \varphi(x_i) + b + e_i \quad (i = 1, 2, \dots, k)$$

引进拉格朗日乘子 α_i , 拉格朗日函数为

$$L(w, b, \rho, \alpha) = J(w, \rho) + \sum_{i=1}^k \alpha_i [w^T \varphi(x_i) + b + e_i - y_i] \quad (i = 1, 2, \dots, k) \quad (11)$$

选择满足 Mercer 条件的核函数:

$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) \quad (i, j = 1, 2, \dots, k) \quad (12)$$

根据 Karush-Kuhn-Tucker (KKT) 条件, 得如下线性方程组:

$$\begin{bmatrix} 0 & 1 \\ 1 & K + \gamma I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (13)$$

其中 K 为由(7)式计算的核矩阵, $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_k)$ 。

解线性方程组, 得 b, α 。使得 $\alpha_i \neq 0$ 的为支持向量。

$$\begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & K + \gamma I \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (14)$$

3.2 支持向量机算法

算法 2

Step1 给定训练集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)\} \in (X \times Y)$, 其中 $x_i \in X \rightarrow \mathbf{R}^n, y_i \in Y \rightarrow \mathbf{R}, i = 1, \dots, k$;

Step2 选择适当的核 $K(x, x)$;

Step3 用(13)、(14)式求解 b, α ;

Step4 支持向量机为

$$y(x) = \sum_{i=1}^k \alpha_i K(x_i, x) + b \quad (15)$$

4 供应商评价的应用

重庆市某中型摩托车制造企业, 为了降低经营风险, 提高企业的经济效益, 拟应用近 1.5 年的历史数据对筛选出的 247 家产品质量合格的供应商就经济合作和效能进行评价, 筛选下一年度重点合作的供应商。

(1) 从数据仓库清洗出相关的评价指标。从供应关系数据仓库中按季度加工清洗出与供应商经济合作和效能评价的相关指标 27 项 (见表 1) 构成向量 $X = (x_1, \dots, x_{27})$ 。

(2) 定义供应商经济合作和效能指数。按照算法 1 将 $X \mapsto X' (\mathbf{R}^n \mapsto \mathbf{R}^n)$, 定义供应商经济合作和效能指数 $Y = \|X'\|$ 。

(3) 优化供应商经济合作和效能评价指标。按照 2.2 的公式计算 X' 与 Y 的偏相关系数 r_{iy} , 从中选择 r_{iy} 值较大的 11 项为供应商经济合作和效能评价指标, 构成向量 $X_{opt} = (x_{opt1}, \dots, x_{opt11})$ 。

表 1 供应商经济合作和效能评价指标及优化

| 类别 | 指标(X) | | r_{ij} | X_{opt} | |
|-------------------|------------------------|----------|-------------|-----------|-------------|
| 供应商基 本信息 | 供应部件种数 | x_1 | 0.000 529 | F | |
| | 供应部件种数/全部部件种数 | x_2 | 0.016 235 | F | |
| 合同信息 | 合同数 | x_3 | 0.011 325 | F | |
| | 合同金额 | x_4 | 0.161 686 | T | x_{opt1} |
| | 合同部件件数/全部供应商合同部件件数 | x_5 | 0.243 516 | T | x_{opt2} |
| | 履行合同时提出的合同变更次数/该供应商合同数 | x_6 | 0.007 592 | F | |
| | 变更合同请求被响应数/提出的合同变更次数 | x_7 | 0.018 952 | F | |
| | 正常履行合同次数/该供应商合同数 | x_8 | 0.265 056 | T | x_{opt3} |
| | 正常履行合同金额/该供应商合同金额 | x_9 | 0.317 294 | T | x_{opt4} |
| | 正常履行合同的到货件数/该供应商合同部件数 | x_{10} | 0.327 265 | T | x_{opt5} |
| | 逾期履行合同次数/该供应商合同数 | x_{11} | 0.094 145 | F | |
| | 逾期履行合同金额/该供应商合同金额 | x_{12} | 0.083 350 | F | |
| | 逾期履行合同的到货件数/该供应商合同部件数 | x_{13} | 0.091 176 | F | |
| 供应(包括运输 和仓储)信息 | 部分履行合同次数/该供应商合同数 | x_{14} | 0.041 456 | F | |
| | 部分履行合同金额/该供应商合同金额 | x_{15} | 0.072 487 | F | |
| | 部分履行合同的到货件数/该供应商合同部件数 | x_{16} | 0.090 648 | F | |
| | 完全未履行合同次数/该供应商合同数 | x_{17} | 0.032 835 | F | |
| | 完全未履行合同金额/该供应商合同金额 | x_{18} | 0.067 162 | F | |
| | 完全未履行合同的到货件数/该供应商合同部件数 | x_{19} | 0.093 272 | F | |
| | 运输、仓储、转运成本/实际购买成本 | x_{20} | 0.116 867 | T | x_{opt6} |
| | 合同购买单价-实际购买单价 | x_{21} | 0.141 372 | T | x_{opt7} |
| 货款支付信息 | 货款总金额 | x_{22} | 0.167 594 | T | x_{opt8} |
| | 货款/全部供应商货款 | x_{23} | 0.098 828 | F | |
| | 货款/同种部件货款 | x_{24} | 0.095 335 | F | |
| 效能信息 | 供应商产生的效益总数 | x_{25} | 0.276 243 5 | T | x_{opt9} |
| | 供应商产生的效益比例 | x_{26} | 0.326 874 5 | T | x_{opt10} |
| | 供应商产生的效益总数/购买该供应商的部件成本 | x_{27} | 0.308 281 0 | T | x_{opt11} |
| 目标值 | 供应商经济合作和效能指数 | y | 1.000 00 | | |

(4)建立供应商经济合作和效能评价支持向量机模型。从数据仓库中钻取出这 247 家产品质量合格的供应商不同批次供货数据 1 285 个,构成训练集 $T = \{(x_{1opt}, y_1), \dots, (x_{1285opt}, y_{1285})\}$,其中 $x_{iopt} \in X_{opt} \rightarrow \mathbf{R}^{11}$, $y_i \in Y \rightarrow \mathbf{R}$, $i = 1, \dots, 1285$,设置精度参数 $\varepsilon = 0.01$,选择高斯核 $K(x, x') = \exp(-\|x - x'\|^2 / \sigma^2)$, $x \in X_{opt} \rightarrow \mathbf{R}^{11}$, $\sigma = 0.5$ 。

按照算法 2 求得 $\alpha = (\alpha_1, \alpha_1^*, \dots, \alpha_1, \alpha_1^*)^T$ 。选择 α 的正分量 $\alpha_j = \alpha_{17} = 0.437 526$,计算 $b = y_j - \sum_{i=1}^l (\alpha_i^* - \alpha_i)K(x_i, x_j) - \varepsilon = 0.067 852$ 。

供应商经济合作和效能评价的支持向量机模型为 $y = \sum_{i=1}^l (\alpha_i^* - \alpha_i)K(x_i, x) + 0.067 852$ (16)

(5)供应商经济合作和效能评价分析。从表 1 看出,供应商经济合作和效能评价指标 27 项,优选 14 项建立支持向量机^[9]模型,大大提高模型的使用效率;支持向量机模型的复相关系数 $\rho = \sqrt{(\hat{y} - \bar{y})(\hat{y} - \bar{y})' / ((y - \bar{y})(y - \bar{y})')} = 0.972 865$;支持向量 15 个,有较强的泛化性。

5 结论

供应商评价是供应链管理的首要问题。供应商评价需要供应商的相关信息,信息如何获取?只要不是新建企业,企业都普遍采用供应商的历史数据去评价供应商,建立供应关系数据仓库可以很好地解决供应商评价的数据问题。尤其是新建的企业,更

应该建立供应关系数据仓库,以便将来更好地评价供应商。供应关系数据仓库已成为了供应商评价的基础。

在供应商评价中,评价的目标变量,例如供应商的合同信用度、供应商的质量指数、企业与供应商经济合作和效能指数等,都没有确切的定义,并且想确切地定义它非常困难,但这些目标变量就像冷热是一个相对大小的值,本文的“温度计-洋葱头”算法提供解决这一难题的简单方法。

人们总试图全面地评价供应商,但事实上,人们只关心供应商某方面,只需从关心的角度去评价供应商,真正能包罗万象的所谓“全面的供应商评价”是没有的,也是没有价值的。选择相关评价指标,并按2.2的方法优化,剔除相关性较弱的指标,可以降低评价和使用评价模型的复杂度。

支持向量机具有很强的泛化性和解决非线性问题的能力,可以用于建立供应商评价的支持向量机模型。特别地,在数据挖掘中,用难于获取信息定义目标变量,用容易获取的信息与目标变量建立支持向量机模型,可以提高建立和使用模型的效率。

参考文献:

[1] PRAHINSKI C, BENTON W C. Supplier Evaluations: Communication Strategies to Improve Supplier Performance[J]. Journal of Operations Management, 2004, 22(1): 39-62.

- [2] HUMPHREYS P K, LI W L, CHAN L Y. The Impact of Supplier Development on Buyer-supplier Performance[J]. Omega, 2004, 32(2): 131-143.
- [3] SCHMITZ J, PLATTS K W. Roles of Supplier Performance Measurement: Indication from a Study in the Automotive Industry[J]. Management Decision, 2003, 41(8): 711-721.
- [4] SIMPSON P M, SIGUAW J A, WHITE S C. Measuring the Performance of Suppliers: An Analysis of Evaluation Processes[J]. Journal of Supply Chain Management, 2002, 38(1): 29-41.
- [5] TALLURI S, NARASIMHAN R. A Note on “A Methodology for Supply Base Optimization”[J]. IEEE Transactions on Engineering Management, 2005, 52(1): 130-139.
- [6] TALLURI S, NARASIMHAN R. A Methodology for Strategic Sourcing[J]. European Journal of Operational Research, 2004, 154(1): 236-250.
- [7] YANG W Q, LIU L T, LIN H Z. Multivariate Statistical Analysis[M]. Beijing: Higher Education Press, 1991.
- [8] CRISTIANINI N, TAYLOR J S. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods[M]. Cambridge: Cambridge University Press, 2000.
- [9] 陈代江. 汇率预报的非线性组合建模与预测方法研究[J]. 重庆师范大学学报(自然科学版), 2003, 20(3): 1-4.

(责任编辑 游中胜)