

基于遗传算法的同步优化方法在财务困境预警中的应用

宋新平^{1,2}, 丁永生^{2,3,4}, 曾月明²

(1. 江苏大学 工商管理学院, 江苏 镇江 212013; 2. 东华大学 旭日工商管理学院, 上海 200051; 3. 东华大学 信息科学与技术学院, 上海 201620; 4. 东华大学 数字化纺织服装技术教育部工程研究中心, 上海 201620)

摘 要: 传统的基于统计技术的变量筛选法不能保证财务困境预警模型的精度。本文提出了用遗传算法同时优化输入变量和支持向量机参数的方法。实证研究表明: 该同步方法在降低变量维数的同时得到较好的预测精度, 其得到的优化变量集也具有较强的经济含义。

关键词: 同步优化; 输入变量; 支持向量机参数; 财务困境预警

中图分类号: F275 文献标识码: A 文章编号: 1003-5192(2009)01-0048-08

Application of A GA-Based Simultaneous Optimization Method in Financial Crisis Prediction

SONG Xin-Ping², DING Yong-sheng^{3,4}, ZENG Yue-ming²

(1. School of Business and Management, Jiangsu University, Zhenjiang 212013, China; 2. Glorious Sun School of Business and Management, Donghua University, Shanghai 200051, China; 3. College of Information Sciences and Technology, Donghua University, Shanghai 201620, China; 4. Engineering Research Center of Digitalized Textile & Fashion Technology, Ministry of Education, Donghua University, Shanghai 201620, China)

Abstract: The variable selection method based on traditional statistical technique can not achieve high accuracy in financial distress prediction. Hence, a method based on genetic algorithm is proposed to optimize the input variables and parameters of support vector machine (SVM) simultaneously. The empirical result indicates that this proposed method can reduce the number of variables and achieve high prediction accuracy. Moreover, the variable subset extracted via this proposed method can be interpreted economically.

Key words: simultaneous optimization; input variable; parameters of support vector machine; financial distress prediction

1 引言

企业陷入财务困境最终影响企业生存, 并且要求利益相关者、社会和国家经济付出高额成本。因此, 建立预警模型有效预测财务困境, 对于投资决策、财务风险诊断、信用风险防范以及管理层监管都具有重要意义。

目前为止, 国外对财务困境及破产的理论和实证研究较为全面和深入。而由于国内资本市场特殊性等诸多因素影响, 对财务困境预测的研究与国外比有较大差距。财务困境的定量预警方法是财务困境预测研究的重要研究方面。多年来, 研究者一直

致力于寻找具有良好预测能力的方法, 包括从最早的统计模型(多元判别分析, 逻辑回归模型等)到专家系统、神经网络、决策树、支持向量机等智能方法。基于这些方法的预警的本质是用分类器对财务困境和财务健康公司进行区分决策, 或者对财务困境程度不同的公司判定成不同类别。从模式识别的角度看, 基于分类原理的财务困境预测模型的精度不仅取决于分类器的性能, 同时也取决于作为模型输入的变量指标。目前, 已有不少国外破产预测研究就分类器的性能取得共识, 一致认为支持向量机的分类精度不亚于或者甚至优于神经网络、决策树、案例推理等智能方法和统计方法^[1]。然而, 由于财

收稿日期: 2007-10-14

基金项目: 教育部哲学社会科学研究重大课题攻关资助项目(07JJD0020); 教育部新世纪优秀人才支持计划资助项目(NCET-04-415); 上海市教育委员会科研创新资助项目(08ZS33)

务困境的经济影响因素复杂众多,迄今尚缺乏财务困境预警指标选择的统一理论。不同研究者采用不同的变量组合,预测结果缺乏可比性,影响了预测模型的可靠度和可信度。诸多预警模型指标的选取都存在明显差异,无法对比哪些指标在模型中更有效;即使同一类型的指标,不同的研究者选取的指标差异也很大。

国外变量选择方法多用单变量显著检验法,方差分析、逻辑回归、逐步判别分析等线性回归法^[2-6]。也有部分文献直接根据经验定性分析确定变量,近年,智能方法被尝试用于变量的提取。Ko利用粒子群等进化算法筛选关键变量,然后输入多元判别分析和BP神经网络分类器。实验结果发现基于进化算法所筛选的变量的预测精度好于逐步回归法^[7]。Lee用粗糙集对输入变量集简约,认为它不仅是有效的神经网络数据预处理器,而且具有降维、省时、防止神经网络过拟和的优点。基于神经网络和粗糙集的复合模型精度优于多元判别分析和神经网络^[8]。

陈静在单变量二分类判定分析中,发现在资产负债率、流动比率、总资产收益率和净资产收益率等四个财务指标中,流动比率与资产负债率的误判率最低^[9]。吴世农从反映盈利、偿债、营运、成长能力的21个财务指标中,采用线性逐步回归法选取盈利增长指数、资产报酬率、流动比率、长期负债与股东权益比率、营运资本与总资产比、资产周转率等6个指标作为多元判定分析的变量^[10]。刘洪通过单变量显著检验(T检验和U检验)从36个备选指标中筛选出22个指标,再剔除相关性指标,最后确定10个指标作为自变量输入BP神经网络^[11]。李波根据前人经验直接选用10个财务指标输入支持向量机进行预测^[12]。杨毓使用了逐步logit回归分析选择显著变量,将财务变量减少到13个作为支持向量机的输入,构建了破产预测模型^[13]。

上述实证研究一般根据财务理论和经验确定预警指标,直接将其作为输入变量(特征);或者统计方法筛选出差异显著指标作为输入变量。这种方法选出的特征,具有较强统计解释性。但是大多数情况下,仅对每个单独特征按照统计或者可分性判断进行排队,取前面D个特征所构成的特征集并非最优特征集,尤其是特征和分类器息息相关的时

候^[14]。另一方面,线性回归法存在局限性,进行的是变量的线性搜索,对特征提取顺序的依赖性导致最终提取的变量组不同。Altman进行无数次的变量组合才得到著名的Z-score模型^[2]。因此,必须开展大量变量组合试验才能保证模型的预测精度。基于此,本文提出了遗传算法同步优化特征和支持向量机参数的方法用于财务困境预测,遗传算法不仅用于从巨大的特征变量组合空间中搜索较优特征组合,也用来寻找较优的支持向量机参数,以此确保预测精度。本文还对比了同步优化方法和统计方法提取的变量组对预测模型的影响。最后通过不同的变量筛选方法识别出财务困境的关键决策变量。

2 遗传算法同步优化变量和支持向量机模型参数的原理

2.1 研究思路

分类过程依赖于待识别对象的一组基本特征。特征选择的目的是辨识关键特征,用较少变量达到精确决策。企业陷入财务困境前会表现出盈利下降、现金流恶化等征兆,从而表现出财务指标的异常,这些指标即可作为甄别财务困境和财务正常企业的原始特征。财务困境预测建模中,通常使用统计方法预选变量再输入分类器,这种过滤式方法可以排除非关键性特征,缩小搜索规模,但是不能保证选出针对特定分类器的优化特征子集。为了获得优化特征集应该把所有特征组合计算出并评价。若D个特征中选择出d个最优或次优特征,组合数是 $N=C_D^d=D!/[(D-d)! \times d!]$ 。实际中需要优化的特征数未知,则所有特征组合数是 2^D 次。财务指标数量多在几十个,组合试验费时费力。

遗传算法是一种基于自然选择机理的随机搜索算法。具有隐并行性和良好的全局寻优能力;自适应地调整搜索方向。这些性质使遗传算法适合解决财务困境特征选择的组合优化问题,使其快速有效地搜索巨大的特征组合空间发现较优特征组。

支持向量机是Vapnik开发出的基于统计原理的机器学习方法,优点是执行结构风险最小化原则,避免了BP神经网络的过拟和、推广能力差和易于陷入局部最优的缺点。目前,支持向量机因其良好的性能而广泛应用。Shin在破产研究发现支持向量机在小样本时依然保持较好的预测性能和推广能力,而神经

网络性能明显下降^[4]。我国资本市场起步晚,数据量小且人为因素干扰大,加之财务变量之间存在复杂的非线性关系。而支持向量机在解决小样本、高维和非线性问题方面具有独特优势,非常适合作为财务困境预测的分类器。

基于上述考虑,本文设计了遗传算法和支持向量机相结合的嵌入式特征选择方法,将其应用于财务困境预测。这种算法是在应用遗传算法筛选特征的过程中直接训练支持向量机,根据支持向量机在验证集上的分类性能评价选出的特征组。但是还有两个重要的方面需要考虑:(1)支持向量机的模型参数选择对其性能的优劣有重要的影响;(2)分类模型的参数与所输入变量组是相互影响^[15]。因此,采用基于遗传算法同时优化支持向量机参数和输入变量组。

2 2 基于遗传算法的特征和支持向量机参数同步优化

2 2 1 支持向量机

对于非线性可分样本,支持向量机构造一个非线性映射 $(\phi \in \mathbb{R}^p \rightarrow \mathbb{R}^{p_k})$ 将样本从 p 维空间映射到更高维特征空间,使该样本在高维空间线性可分。若用 $Y=[y_1, \dots, y_m]^T$ 来表示 m 个样本的类标号, $y_i \in \{-1, +1\}$ 表示财务困境和财务正常两类,用 $x_i=(x_i^1, \dots, x_i^p) \in \mathbb{R}^p$, $i=1, \dots, m$ 表示 m 个样本的 p 个变量值。应用中每个样本都可以用一个超平面进行划分,该优化超平面满足将所有的训练集区分开的要求,要寻找的最优超平面实际上等价于求一个二次规划问题^[16]。如(1)式所示

$$\min_{w, b, \xi} \left(\frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i \right) \quad (1)$$

$$s.t. \begin{cases} y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i & i=1, \dots, N \\ \xi_i \geq 0 \end{cases}$$

求解得到非线性可分条件下两分类的最优超平面,如(2)式所示

$$f(x) = \text{sgn} \left(\sum_{i=1}^N \alpha_i y_i K(x, x_i) + b \right) \quad (2)$$

其中 α_i 表示拉格朗日系数, b 表示超平面的偏置项,可以通过求解凸二次规划问题得到^[15]。 $K(x, x_i) = \phi(x)^T \phi(x_i)$ 是满足 Mercer 条件的核函数。高斯核函数 $k(x, x_i) = \exp(-\gamma \|x - x_i\|^2)$ 的稳定和推广性能较好,并且需要优化的参数少,所以本文使用高斯核支持向量机作为嵌入式特征选择方法的分类器。RBF核函数的支持向量机有两个主要参数即核参数 γ 和惩罚因子 C 。核参数影响样本数

据在高维特征空间中分布的复杂程度,而误差惩罚因子 C 的作用是调节学习机的置信范围和经验风险的比例。核参数和惩罚因子选择不当会导致模型的过拟和或拟和不足,从而导致模型性能下降。

2 2 2 遗传算法同步优化的模型设计

本文将遗传算法和支持向量机有机结合,利用遗传算法的杂交和变异选择算子在全变量空间搜索变量组合解和高斯核支持向量机模型参数。重点考虑以下三方面。

(1) 遗传算法编码

遗传算法是在由码串表示的个体所组成的群体上进行遗传算子的操作。基于上述分析,遗传算法需要同时优化两个方面:财务困境预测模型的输入变量,模型参数 γ 和惩罚因子 C 。所以,每个染色体代表了需要优化的特征子集、 C 和 γ 。本文采用二值编码方法对个体编码。个体码串由三部分组成。第一部分是特征掩码,采用二值编码,其中每一位分别与某给定特征所对应,取值“1”表示特征子集中有该特征,否则表示没有该特征。第二部分和第三部分分别代表 C 和 γ ,其编码长度可以根据精度需要(3)式进行调整。

$$l = \lceil \log_2 \frac{\max_b - \min_b}{2^{-1}} \rceil \times d \quad (3)$$

其中 l 是二进制码串, \max 和 \min 分别是参数的上下界, $\lceil \cdot \rceil$ 是二进制码串的长度, d 是一个二进制代表的精度。

如图1所示的码串个体,表示10个变量中的第1、5、6、7、8、9的变量构成了特征子集,在此特征集下的高斯核支持向量机参数对 (C, γ) 。本混合编码方法可以对表示特征子集的二值码串进行标准的遗传算子运算,从而保留了典型遗传算法方法的优点,计算简单,遗传算子对大规模特征组合空间的搜索非常有效,易于扩展到大规模特征集寻优。起始的群体中的起始码串采用均匀分布的随机分布函数产生,并成为下一次遗传搜索过程中个体的父代。

特征掩码编码	C编码	γ编码
1000111110	1000000110	1000111110

图1 编码

(2) 遗传算法的适应度函数

用来评价一组所选出的特征的依据有:①用该特征组训练得到的分类器在验证集上的性能指标;②特征数量的多少,它直接影响数据获得的成本和

计算效率;③是否具有可解释的实际意义。由于目前我国财务困境数据量小,特征数量的影响忽略不计,并且本文主要关心导致分类器性能最优的特征组合。因此,采用国际上通用的交叉验证法来评估特征组在训练集上的性能,在测试集上使用分类器的准确度作为评价指标。适应度函数采用训练集的10折交叉验证法的准确度,这种方法有助于得到优化的支持向量机模型参数,防止模型出现过拟和现象^[17]。具体而言,是将训练样本随机分成 n 份,用第一份作为训练子集训练支持向量机,用训练过分分类器对其余 $n-1$ 份测试,从而得到第一个验证集准确度,轮流依次计算其他验证集准确度,最后计算这 n 个验证集准确度的平均值就是适应度。本文选用10折交叉验证法的准确度作为适应度函数。

(3)遗传算法的操作算子

本文采用轮盘选择算子选取进行下一次遗传运算的父母码串。对码串进行标准遗传算法的交换算子的运算,采用两点交叉法。对码串进行标准遗传算法的变异算子的运算,采用位变异。同时,程序提供了修正机制对无效特征子集进行规避处理,如没有选中人和特征的空集。

(4)同步优化算法流程设计

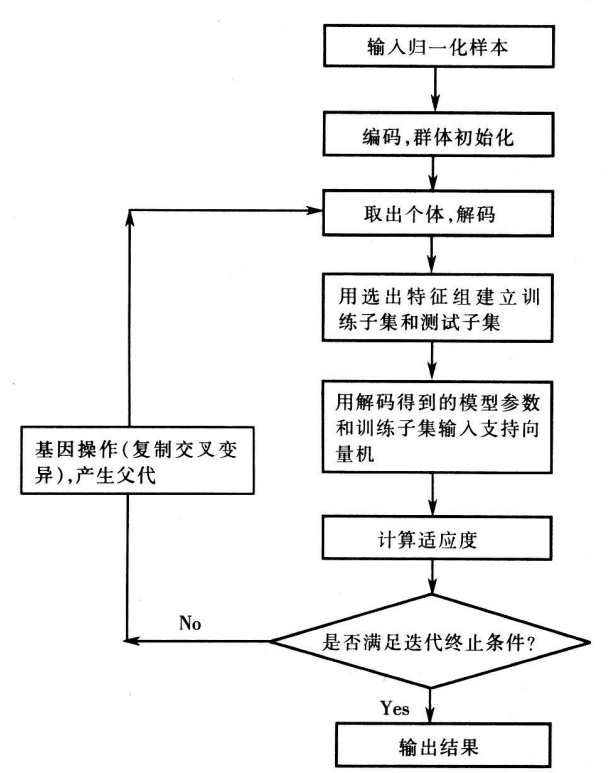


图 2 算法流程图

算法流程如图 2所示。遗传算法可以看作是一个由可行解组成的群体逐代进化的过程。算法步骤如下:①按照一定精度要求在规定区间内随机产生二进制编码,即得到初始染色体群;②将染色体解码得到相应的特征子集和模型参数 γ 和 C 建立基于此特征子集的训练集和测试集,建立基于此模型参数的高斯核支持向量机分类器,求解该群体的适应度;③依据适应度和设定的交叉率、变异率等参数对染色体群进行选择,交叉,变异等遗传操作,获得父代群体;④若进化达到设定的代数或者精度要求则算法停止,否则转步骤 2⑤输出优化的特征子集和模型参数。

3 同步优化模型在财务困境预警中的应用

3.1 预警指标体系建立

根据财务理论和国内外研究结果,财务困境前会出现一些警兆,并主要表现为赢利能力、流动性和偿债能力、营运效率、成长能力、现金流等财务指标的变化,这些指标覆盖企业生产经营的全过程。根据实际和数据可得性,本文纳入尽可能多的预警指标,并试图通过同步优化方法对这些指标进行客观选择。

3.2 样本准备

国内多数研究以特别处理(ST)作为财务困境的界定标准,因为它具有容易观察的优点。本文结合上市公司实际,以财务状况异常的ST公司作为财务困境样本,而以其他非ST公司作为非财务困境样本。由于行业、技术类型、规模时间等因素会对预测模型产生影响^[18],把研究对象限制在信息、生物医药、材料和机械等科技制造行业的上市公司。规模用资产总计表示,用Log处理。为避免会计准则调整对财务指标影响,限于2002~2006年度。剔除审计意见为拒绝和保留的样本,同时剔除其他数据缺失和异常、数据操纵的可疑样本。采用ST事件前的数据构建预测模型。对于ST样本, t 代表发生ST事件的年度, $t-1$ 代表ST事件前的一个年度,用 $t-1$ 年度的数据预测 t 年是否为财务困境公司。由于某个年度发生ST事件的样本数量很少,故以2001~2006年度期间发生ST事件的公司为财务困境对象。把2005年所有非ST公司作为非财务困境样本,以2004年的数据预测2005年的情况。增长率指标指 t 年相对于 $t-1$ 年的的百分比率。

按照上述原则从深沪两市A股市场取样。发

现 ST公司中可疑样本较多,排除这些可疑样本后使得原本很少的 ST公司数量更少。因此,财务困境样本取自 ST事件的前一年和前两年。最终样本包含财务困境样本 56个,非财务困境样本 194个。样本

分为训练集 125个和测试集 125个。随机划分总样本,得到的训练集和测试集分别包含 28个财务困境样本和 97个非财务困境样本。为了减少预测误差,本文对样本数据进行了归一化处理。

表 1 预警指标

	预警指标
偿债能力	营运资本资产率 (X1)
	流动比率 (X2), 速动比率 (X3), 现金比率 (X4), 流动负债经营活动净现金流比 (X5), 资产负债率 (X6), 负债权益比率 (X7), 股东权益比率 (X8), 有形资产债务率 (X9), 债务与有形净值比率 (X10)
营运能力	应收帐款周转率 (X11), 存货周转率 (X12), 流动资产周转率 (X13), 总资产周转率 (X14), 固定资产周转率 (X15), 长期资产适合率 (X16), 负债结构比率 (X17), 长期负债权益比率 (X18)
赢利能力	主营业务利润率 (X19), 成本费用利润率 (X20), 净资产收益率 (X21), 每股净收益 (X22), 总资产利润率 (X23), 主营业务比率 (X24)
成长能力	主营业务收入增长率 (X25), 应收帐款增长率 (X26), 净利润增长率 (X27), 总资产扩张率 (X28), 每股收益增长率 (X29)
现金流	每股营业现金流量 (X30), 现金自给率 (X31), 经营活动产生现金净流量增长率 (X32)
规模	资产总计 (Log) (X33)

3.3 仿真实验

本文中,通过大量实验对遗传算法的相关参数进行试算,得到适应度高的遗传算法参数为:群体数量 100 交叉概率 0.7 变异概率 0.02 特征掩码的长度是 32 高斯核支持向量机的参数 γ 和 C 根据精度要求实验调整,位数在 20~40 遗传算法开始时,在变量组合空间和支持向量核参数的上下界区间内随机产生若干个体。随着群体的不断进化,目标函数值不断增加,新的群体进化出一些新的优良性质。迭代终止条件是:最近 100代适应度不增加。在 360代时满足迭代终止条件,适应度达到 93.2%,优化的特征子集和参数结果见表 2 和表 3 进化过程见图 3 同步优化方法使用了 LBSVM核心模块函数用 VC++开发的。

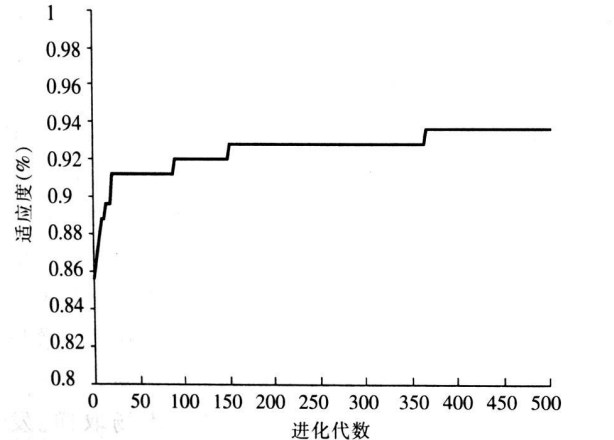


图 3 遗传算法优化特征和模型参数的进化过程

4 同步优化模型的比较与分析

4.1 预测性能的比较

为了说明本文所设计的同步优化方法的有效性,将本方法与基于统计方法的过滤式特征选择方法做一比较。过滤式方法的实现流程是先用统计方法选出特征,然后将其输入高斯核支持向量机,经过参数寻优、模型训练学习、测试等步骤得到预测结果。由于逻辑回归是金融应用中性能较好的统计分类法,以其作为比较基准,同时观察不同特征选择方法对其影响。

本文采用财务困境预测中最常见的筛选特征的统计方法,包括 Anova方差法、TW检验法、逐步判别分析和逐步逻辑回归。这些方法可以判定哪些变量具有对财务困境组和非财务困境组的判别效力,选中的变量输入支持向量机。Anova法指用 Anova单因素方差分析选出组之间有显著差异的变量 ($P < 0.05$)。TW选择特征的方法是:对符合正态分布的变量做 T检验和不符合正态分布的变量做 Wilcoxon非参数检验,选出组之间有显著差异的 ($TW < 0.05$)。逐步判别分析采用前向法,一边判别一边引进组间判别的贡献最大的变量。逐步逻辑回归采用前向法,过程与逐步判别分析类似。限于篇幅,统计方法选择特征的细节不再赘述。这些方法是在 SPSS 2上完成的。

为了保证与同步优化方法的可比性和一致性,

过滤式特征选择方法也使用高斯核支持向量机, 利用了 LIBSVM 进行模型参数的寻优, 凭借网格搜索技术和交叉验证技术进行核参数寻优^[17]。这种参数寻优方法既可以并行运算, 也可以减少过拟和。过滤式方法的步骤是: (1) 用 10 折交叉验证法作为评估不同参数对下分类器性能优劣的依据, 在 ($C=2^{-5}, 2^{-3}, \dots, 2^{15}; \gamma=2^{-15}, 2^{-13}, \dots, 2^3$) 组

合的网格内穷举搜索每个点, 搜索到交叉验证准确度最高的点即为最优参数对 (C, γ); (2) 利用训练集训练基于此最优参数对的分类器, 获得支持向量和 Lagrange 系数等模型参数; (3) 利用训练过的模型对测试样本分类。不同特征提取方法的结果如表 2 所示。分析如下:

表 2 不同特征提取方法的结果比较

方法		特征数量	逻辑回归		支持向量机					
			训练集准 确度（%）	测试集准 确度（%）	训练集准 确度（%）	测试集准 确度（%）	最优参数对		交叉验证 准确度（%）	支持向 量数
							C	γ		
所有特征		33	100	69.6	95.2	83.2	128	0.03125	87.2	45
统计 方法	Anova方差分析	17	88	72.8	90.4	75.2	32	0.125	84.8	50
	TW显著检验	14	84.8	77.6	99.2	77.6	32	8	83.2	61
	逐步逻辑回归	10	89.6	78.4	99.2	80.8	32678	0.125	89.6	30
	逐步判别分析	4	86.4	66.4	97.6	72	8	2	91.2	51
同步优化方法		12			96.8	84	1.839	196.1	93.6	40

表 3 不同方法提取特征子集的比较

变量 编号	全特征	统计方法				同步 方法	选中 频次	变量 编号	所有 特征	统计方法				同步 方法	选中 频次
		Anova	T-W	逐步判 别分析	逐步 Logit 回归					Anova	T-W	逐步判 别分析	逐步 Logit 回归		
		32	17	14	4	10	12			32	17	14	4	10	12
X1							0	X17							0
X2							0	X18							1
X3							2	X19							1
X4							2	X20							1
X5							3	X21							2
X6							2	X22							4
X7							1	X23							4
X8							2	X24							
X9							2	X25							1
X10							1	X26							
X11							3	X27							1
X12							3	X28							
X13							5	X29							1
X14							5	X30							2
X15							1	X31							3
X16							0	X32							0
								X33							3

(1) 不同方法预选的特征集各不相同。逐步判别分析选出的特征数量最少, Anova 法预选出的特征集较多, 而同步优化方法预选出的特征数量适中。

(2) 就过滤式特征选择和支持向量机组合方法而言, 不同的特征集输入支持向量机, 最优模型参数对、交叉验证准确度、支持向量数量都有所不

同, 导致最终测试集的预测结果不同, 这证实了特征集和分类器相互影响的观点^[15]。

(3) 就特征选择和支持向量机组合方法而言, 同步优化方法的准确度最高, 略高于全特征集的预测结果。值得关注的是: 金融财务领域中常用的实证惯例是用统计方法预选变量, 但本实验结果表明基于全特征集的分类准确度未必低于基于统计方法预选特征的准确度; 并且在样本中等规模时, 二者的时间耗费差异不大。基于不同统计方法预选特征的准确度有所不同, 逐步逻辑回归的较高, 逐步判别分析的较低, Anova方差分析和 TW显著检验的准确度介于二者之间。

(4) 同一特征子集下, 支持向量机的预测精度高于逻辑回归。不同特征子集下, 支持向量机和传统逻辑回归法没有可比性, 这说明以往不同文献采用不同变量组合的预测精度结果缺乏可比性。逻辑回归方法如选择合适的变量, 精度可能高于支持向量机。

(5) 从预测性能和变量数量看, 同步优化方法的准确度最高且变量数目适中, 优于其他方法。但是由于同步优化是一个复杂的组合搜索问题, 比较耗费时间。值得注意的是, 单个的基于统计过滤式途径的特征选择方法虽然时间耗费少, 但不能保证选出较优的特征子集, 印证了部分学者的观点^[14]。如为保证统计过滤式的特征选择方法选出预测精度较高的特征子集, 须进行大量变量组合实验。从这个意义看, 同步优化方法是一个确保财务困境预测性能和成本的较优策略。

4 2 关键变量识别与分析

从表 3 可以看出: (1) 同步方法选出了 12 个特征变量, 其中的 11 个和统计方法选出的一样, 说明该方法与其他方法有较好一致性, 选中的变量也有一定的统计解释意义, 符合良好的特征选择方法的标准。(2) 不同方法预选出的特征集不同, 如果不同方法选中的频次高, 可以说明这个变量对于财务困境决策比较重要。频次超过 1 次的有 24 个财务指标, 基本覆盖了企业资本经营周转全过程的财务赢利、偿债能力、成长能力、营运能力、现金流等方面, 其对财务困境预警的贡献度较大。频次 5 次的有流动资产周转率 (X₃) 和总资产周转率 (X₄)。频次 4 次的有每股净收益 (X₂₂) 和总资产利润率 (X₂₃)。频次 3 次的有: 流动负债经营活动净现金流比 (X₅)、应收帐款周转率 (X₁₁)、存货周转率 (X₁₂)、现金自给率 (X₃₁)、资产总计

(X₃)。频次 2 次的有 7 个指标。其中, 每股营业现金流量、现金自给率和流动负债经营活动净现金流比是现金流量指标, 现金自给率指通过经营活动创造的现金净流量能否满足资本支出、存货投资支出和支付现金股利的需要; 每股营业现金流量反映了每股盈利的支付保障的现金流量; 流动负债经营活动净现金流比反映了本期经营活动净现金流量偿还债务能力, 它们揭示了收益质量, 不容易受到经营者的主观歪曲, 在我国资本市场利润操纵的背景下具有现实意义; 每股净收益和总资产利润率反映了财务赢利能力, 而陷入财务困境的企业通常获利能力低, 现金支付能力不足; 规模也对财务困境预警模型有影响, 上述这些指标对财物危机决策具有较重要的作用, 也基本支持马若薇利用信息熵对预测变量的权重排序结果^[15]。应收帐款周转率和存货周转率等反映企业经营状况的指标对于财务困境预警起一定作用。

5 结论

财务困境预测模型研究的常规范式是采取统计方法选出特征, 其本意是为了减少变量数目和提高预测精度, 同时提高运算速度。但是, 本文的实证分析发现: 财务困境预测模型应用中, 传统统计过滤式途径的变量选择法不能保证获得针对支持向量机的最优子集 (预测精度高), 必须结合特定问题和样本数据进行大量变量的组合实验方能确保预测效果, 显然这种作法耗时耗力。为此, 本文提出了遗传算法同步优化特征和分类器参数的方法。实验结果表明同步优化方法在缩减变量数量的同时保持较好的分类精度。随着财务困境样本的增多以及支持向量机在财务困境预测系统中的应用商业化, 同步优化方法是一种较优的时间性能和预测精度折中策略, 具有较高的实用价值。同时, 该方法也可以广泛应用于如企业评级、投资风险评、信用评级、财务舞弊识别、保险欺诈检测等基于分类原理的金融风险管理领域。

对于财务困境预测模型, 财务指标数量众多, 关系复杂, 目前尚缺少财务困境预警指标选择的理论框架。实证发现 4 种不同方法选择的指标不完全相同。通过比较不同方法选出的变量, 客观地发现了财务困境决策的关键变量, 不但给管理人员提供有益的警示信息, 而且也为财务困境预警指标选择研究提供了一个新的视角。

建立上市公司财务困境预警模型和系统是一

个庞大复杂的工作,需要随着资本市场的健全和样本数量的增多不断对预警模型验证和完善。

参 考 文 献:

- [1] Min JH, Lee Y C. Bankruptcy prediction using support vectormachine with optimal choice of kernel function parameters [J]. Expert Systems with Applications, 2005 28: 603-614
- [2] Altman E J. Financial ratios discriminant analysis and the prediction of corporate bankruptcy [J]. The Journal of Finance, 1968 13(4): 589-609
- [3] Beaver W. Financial ratios as predictors of failure [J]. Journal of Accounting Research, 1966 4: 71-111
- [4] Shin K S, Lee T S, Kim H J. An application of support vectormachines in bankruptcy prediction model [J]. Expert Systems with Applications, 2005 28: 127-135
- [5] Lin F Y, McLean S. A data mining approach to the prediction of corporate failure [J]. Knowledge-Based Systems, 2001 14: 189-195
- [6] Lee K D, Booth D, Alan P. A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms [J]. Expert Systems with Applications, 2005 29: 1-16
- [7] Ko P C, Lin P C. An evolution-based approach with modularized evaluations to forecast financial distress [J]. Knowledge-Based Systems, 2006 19: 84-91
- [8] Ahn B S, Cho S S, Kim C Y. The integrated methodology of rough set theory and artificial neural network for business failure prediction [J]. Expert Systems with Applications, 2000 18: 65-74
- [9] 陈静. 上市公司财务恶化预测的实证分析 [J]. 会计研究, 1999 (4): 31-38
- [10] 吴世农, 卢贤义. 我国上市公司财务困境预测模型研究 [J]. 经济研究, 2001 (6): 46-55
- [11] 刘洪, 何光军. 基于人工神经网络方法的上市公司经营失败预警研究 [J]. 会计研究, 2004 (2): 42-46
- [12] 李波, 何建敏. SVM在企业财务困境分析中的应用 [J]. 现代管理科学, 2004 (12): 35-39
- [13] 杨毓, 蒙肖莲. 用支持向量机构建企业破产预测模型 [J]. 金融研究, 2006 (10): 65-75
- [14] 边肇祺, 张学工. 模式识别 [M]. 北京: 清华大学出版社, 2000
- [15] Föhlich H, Chapelle Q. Feature selection for support vector machines by means of genetic algorithms [A]. Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence [C]. California: Sacramento Press, 2003: 142-148
- [16] Vapnik V N K. The nature of statistical learning theory [M]. New York: Springer-Verlag, 1995
- [17] Hsu C W, Chang C C, Lin C J. A practical guide to support vector classification [EB/OL]. <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf> 2008-05-21
- [18] 马若微. 基于粗糙集和信息熵的上市公司财务困境预警指标的确立 [J]. 当代经济科学, 2005 27(2): 45-50
- [12] 刘海潮, 李垣. 战略风险管理——战略管理研究的新前沿 [J]. 管理工程学报, 2003 17(3): 4-7
- [13] 刘益, 李垣, 杜旆丁. 基于资源风险的战略联盟结构模式选择 [J]. 管理科学学报, 2003 6(4): 34-42
- [14] Anderson E, Weitz B. Determinants of continuity in conventional industrial channel dyads [J]. Marketing Science, 1989 8(4): 310-323
- [15] 刘益, 钱丽萍, 尹健. 供应商专项投资与感知的合作风险: 关系发展阶段与控制机制的调节作用研究 [J]. 中国管理科学, 2006 (1): 30-36
- [16] Doney P M, Cannon J P. An examination of the nature of trust in buyer-seller relationships [J]. Journal of Marketing, 1997 61(2): 35-51
- [17] Young Y Bara C, Wiersma M. Strategic flexibility in information technology alliance: the influence of transaction cost economics and social exchange theory [J]. Organization Science, 1999 10(4): 439-460
- [18] Young L C, Wilkison I F. The role of trust and cooperation in marketing channels: a preliminary study [J]. European Journal of Marketing, 1989 23(2): 109-123
- [19] Joshi A W, Arnold S J. The impact of buyer dependence on buyer opportunism in buyer-supplier relationships: the moderating role of relational norms [J]. Psychology & Marketing, 1997 14(18): 823-845
- [20] Haugland S A. Factors influencing the duration of international buyer-seller relationships [J]. Journal of Business Research, 1999 46: 273-280
- [21] Kim K. On interfirm power, channel climate and solidarity in industrial distributor-supplier dyads [J]. Journal of the Academy of Marketing Science, 2000 28(3): 388-405
- [22] Kumar N, Scheer L K, Jan-Benedict E M S. The effects of supplier fairness on vulnerable resellers [J]. Journal of Marketing Research, 1995 32(1): 54-65

(上接第 41 页)