

制造业上市公司财务预警体系的构建及比较

——基于数据挖掘技术

王 艺¹, 姚正海²(教授)

【摘要】 本文以2015年A股制造业上市公司为研究对象,选取反映企业偿债能力、经营能力等六个方面的27个指标构建预警指标体系,并运用T值显著性检验和因子分析法对预警指标进行约简。在此基础上,本文利用Logistic回归、决策树和BP神经网络分别构建财务预警模型,并对各个模型的优缺点及预测结果进行比较分析。研究表明:随着ST发生时间的临近,预警准确率总体呈上升趋势;从中长期来看,决策树和BP神经网络模型的预测能力优于Logistic回归模型;从短期来看,三种模型的预测能力没有显著差异。

【关键词】 财务预警; 因子分析; Logistic回归; 决策树; BP神经网络

【中图分类号】 F224

【文献标识码】 A

【文章编号】 1004-0994(2016)21-0049-7

一、引言

随着国际经济一体化的步伐加快,上市公司在把握改革和发展机遇的同时,也面临着更加激烈的竞争和更大的风险。多数企业往往面临的是生存问题,然后才是发展问题。因此,企业有必要建立财务预警系统,在危机到来之前向企业经营管理者发出警告,把危机消灭在萌芽状态。但是,如何从企业经营管理者产生的大量业务数据中挖掘出对企业财务危机有预警作用的信息呢?这是上市公司亟待解决的一个重要问题。数据挖掘技术的不断发展和成熟运用正好解决了这一问题,它能够从企业大量的业务数据中提取并高度自动化地分析数据,进而帮助决策者做出正确的决策。

本文以制造业上市公司为研究对象,利用传统的Logistic回归方法及数据挖掘技术中常用于预测模型的决策树和BP神经网络分别构建财务预警模型,并对不同预测模型的优缺点以及预测结果进行比较分析,以期对制造业上市公司提供切实可行的财务预警方法。

二、文献综述

(一)财务危机的界定

国外的研究通常将财务危机定义为破产,如Altman(1968)、Tzong-Huei Lin(2009)、Jia-Jane Shuai(2005)。国内对于财务危机的界定绝大多数以被ST为标准。这主要是由于我国特殊的国情,上市公司鲜有破产者,如果以此作为危机的标准,难以获取样本数据。对此,不少学者提出了质疑,吴

兴泽(2011)认为,被特别处理不能等同于财务危机,公司被ST只是公司经营不善的一个信号。本文认为,在现有以财务指标对财务危机进行预警研究的大框架下,使用ST和非ST作为区分财务危机样本与非财务危机样本的标准也未尝不可。因为,无论被ST还是破产,都是企业希望避免的。

(二)财务预警模型研究现状

用于建立财务预警模型的方法可分为两大类:统计分析方法和数据挖掘技术。Altman(1968)最早将多元线性判别方法引入财务风险预警领域创立了Z值模型,预测准确率高达95%。但是,Z值模型要求预警变量符合联合正态分布,而现实中大多数企业的财务指标难以满足这一要求。为克服这一缺陷,以Ohlson(1980)为代表的学者提出采用Logistic回归判别方法来提高财务风险的预警能力。

国内学者陈静(1999)最先致力于财务困境预测的研究,为考察国外的预测模型是否能够有效地应用于国内市场,文章同时建立了一元判定模型和多元线性判定模型。吴世农、卢贤义(2001)分别使用单变量判定模型、多元线性判定模型和Logistic回归分析三种方法建立预测模型,结果发现Logistic回归模型的误判率最低,在前一年仅为6.47%。

随着数据挖掘技术的发展,部分学者开始把知识发现类数据挖掘技术应用到财务预警领域。例如,Wilson和Sharda(1994)运用神经网络模型对意大利企业的财务危机进行判断,准确率高达97%。马若微(2005)运用粗糙集与信息熵原理

【基金项目】 江苏师范大学研究生科研创新项目“基于数据挖掘技术的制造业上市公司财务预警研究”(项目编号:2015YYB146)

□ 改革与发展

分行业建立了上市公司财务困境模型,并将现金流量指标引入预警模型。姚靠华、蒋艳辉(2005)应用决策树技术建立了中国上市公司的财务风险预警系统,结果表明该模型具有较好的预测性。李艳玲、宋喜莲(2014)在Z-SCORE财务预警模型的基础上,利用数据挖掘时序分析方法,对预测结果进行数据分析,结果表明该方法的判断正确率达到80%。

张亮、张玲玲(2015)引入信息融合技术对不同数据挖掘方法得到的结果进行融合处理,发现该方法下的预测准确率要高于单独的支持向量机和Logistic回归两种方法。

综观国内外研究成果,在构建财务预警模型的过程中,还存在以下两方面的问题:①在选取样本上没有划分行业。由于我国市场经济的区域和行业发展并不平衡,各行业的市场竞争程度存在差异,各类企业的会计处理方法和会计政策选择也不同,因此建立财务预警模型应区别行业的类型分别进行研究。②建模方法单一。国内的研究大部分使用的是传统统计分析方法,近年来也有学者致力于将数据挖掘技术引入财务预警领域,但是鲜有学者利用传统统计分析方法与数据挖掘技术同时建模。

为了克服以上缺陷,本文以制造业上市公司为研究对象,利用传统的Logistic回归方法及数据挖掘技术中常用于预测模型的决策树和BP神经网络分别构建财务预警模型,以期为公司提供最优的财务预警方法。

三、研究样本与财务预警指标体系构建

(一)研究样本

1. 数据样本的选取。本文选取2015年A股中制造业上市公司为研究对象。之所以选择制造业,是因为截至2015年12月,全部A股上市公司中制造业上市公司占公司总数的63.2%,据此建立的财务预警模型适用范围较广。本文在原始样本的基础上,剔除因其他状况异常被特别处理的公司和前五年数据缺失的上市公司,最终得到39家被ST公司和922家正常公司。

2. 时间跨度的选择。为了从中长期和短期的角度全面探讨预警模型的预测能力,本文在选择预警时段时,将样本数据的收集时间延伸至公司被ST的前5年,定义上市公司被ST当年为T年。根据我国上市公司信息披露制度,ST公司发生财务危机的实际年份应该是其被ST的前一年。可见,用T-1年的财务数据来预测T年是否发生财务危机没有实践价值。因此,本文最终确定的数据时间跨度为T-6年至T-2年,即利用2009~2013年的数据对公司2015年是否会被特别处理进行预测。

(二)财务预警指标体系的建立

为全面反映企业的财务状况,本文选取了反映企业偿债能力、经营能力、盈利能力、现金流水平、风险水平及发展能力这六个方面的27个指标构建初始预警指标体系,具体如下表1所示。

表1 财务预警指标体系

| 指标类型 | 编号 | 指标名称 | 计算公式 |
|-------|-----------------|----------|--|
| 偿债能力 | X ₁ | 流动比率 | 流动比率/流动负债 |
| | X ₂ | 速动比率 | (流动资产-存货)/流动负债 |
| | X ₃ | 现金比率 | 现金及现金等价物期末余额/流动负债 |
| | X ₄ | 利息保障倍数 | (净利润+所得税费用+财务费用)/财务费用 |
| | X ₅ | 资产负债率 | 负债合计/资产总计 |
| | X ₆ | 产权比率 | 负债合计/所有者权益合计 |
| 经营能力 | X ₇ | 应收账款周转率 | 营业收入/应收账款平均占用额 |
| | X ₈ | 存货周转率 | 营业成本/存货平均占用额 |
| | X ₉ | 流动资产周转率 | 营业收入/流动资产平均占用额 |
| | X ₁₀ | 固定资产周转率 | 营业收入/固定资产平均净额 |
| | X ₁₁ | 总资产周转率 | 营业收入/平均资产总数 |
| 盈利能力 | X ₁₂ | 资产报酬率 | (利润总额+财务费用)/平均资产总额 |
| | X ₁₃ | 总资产净利润率 | 净利润/总资产平均余额 |
| | X ₁₄ | 净资产收益率 | 净利润/股东权益平均余额 |
| | X ₁₅ | 营业净利率 | 净利润/营业收入 |
| 现金流水平 | X ₁₆ | 净利润现金净含量 | 经营活动产生的现金流量净额/净利润 |
| | X ₁₇ | 销售收入现金含量 | 销售商品、提供劳务收到的现金/营业收入 |
| | X ₁₈ | 全部现金回收率 | 经营活动产生的现金流量净额/资产总计 |
| | X ₁₉ | 营运指数 | 经营活动产生的现金流量净额/经营所得现金 |
| 风险水平 | X ₂₀ | 财务杠杆 | 每股收益变化的百分比/息税前利润变化的百分比 |
| | X ₂₁ | 经营杠杆 | 息税前利润变化的百分比/营业收入变化的百分比 |
| | X ₂₂ | 综合杠杆 | 每股收益变化的百分比/营业收入变化的百分比 |
| 发展能力 | X ₂₃ | 总资产增长率 | (资产总计本期期末值-资产总计本期期初值)/资产总计本期期初值 |
| | X ₂₄ | 净利润增长率 | (净利润本年本期单季度金额-净利润上一个单季度金额)/净利润上一个单季度金额 |
| | X ₂₅ | 可持续增长率 | 净资产收益率×收益留存率/(1-分子) |
| | X ₂₆ | 所有者权益增长率 | (所有者权益本期期末值-所有者权益本期期初值)/所有者权益本期期初值 |
| | X ₂₇ | 每股净资产增长率 | (每股净资产本期期末值-每股净资产本期期初值)/每股净资产本期期初值 |

四、制造业上市公司财务预警的实证检验

(一) 财务预警指标体系的约简

1. 制造业上市公司财务预警指标的显著性检验。本文通过 T 值显著性检验, 分析被特别处理的公司与正常公司之间

各年度财务指标的平均值是否存在显著差异, 进而对所选择的财务预测指标进行筛选。T 检验建立的原假设(H_0)为: 被 ST 公司与正常公司在该财务指标上没有差别。各预测年度 T 检验结果如表 2 所示:

表 2 预测指标 T 值检验结果

| 编号 | 指标名称 | T-6(2009年) | | T-5(2010年) | | T-4(2011年) | | T-3(2012年) | | T-2(2013年) | |
|-----------------|----------|------------|-------|------------|-------|------------|-------|------------|-------|------------|-------|
| | | t | Sig. | t | Sig. | t | Sig. | t | Sig. | t | Sig. |
| X ₁ | 流动比率 | -3.044 | 0.002 | -2.525 | 0.012 | -4.431 | 0.000 | -2.703 | 0.007 | -4.092 | 0.000 |
| X ₂ | 速动比率 | -2.724 | 0.007 | -2.486 | 0.013 | -1.881 | 0.060 | -2.630 | 0.009 | -6.345 | 0.000 |
| X ₃ | 现金比率 | -2.342 | 0.019 | -2.231 | 0.026 | -1.556 | 0.120 | -2.237 | 0.025 | -7.197 | 0.000 |
| X ₄ | 利息保障倍数 | -1.516 | 0.130 | -0.320 | 0.750 | -1.062 | 0.293 | -1.347 | 0.178 | -1.450 | 0.151 |
| X ₅ | 资产负债率 | 2.548 | 0.013 | 3.228 | 0.002 | 5.531 | 0.000 | 5.006 | 0.000 | 4.958 | 0.000 |
| X ₆ | 产权比率 | 1.276 | 0.207 | 2.117 | 0.034 | -1.782 | 0.075 | 6.314 | 0.000 | 0.053 | 0.958 |
| X ₇ | 应收账款周转率 | -1.050 | 0.294 | -1.059 | 0.290 | -1.453 | 0.147 | -0.861 | 0.390 | -1.336 | 0.182 |
| X ₈ | 存货周转率 | -0.601 | 0.551 | -0.132 | 0.896 | 1.149 | 0.251 | 5.816 | 0.000 | 6.054 | 0.000 |
| X ₉ | 流动资产周转率 | -0.948 | 0.349 | 0.718 | 0.477 | 0.620 | 0.538 | 0.620 | 0.539 | 0.131 | 0.896 |
| X ₁₀ | 固定资产周转率 | -2.182 | 0.033 | -1.293 | 0.204 | -2.238 | 0.030 | -2.463 | 0.017 | -3.188 | 0.002 |
| X ₁₁ | 总资产周转率 | -3.188 | 0.003 | -2.126 | 0.040 | -2.045 | 0.047 | -1.872 | 0.068 | -2.489 | 0.016 |
| X ₁₂ | 资产报酬率 | -5.562 | 0.000 | -1.093 | 0.280 | -2.680 | 0.010 | 1.972 | 0.049 | -6.647 | 0.000 |
| X ₁₃ | 总资产净利润率 | -6.068 | 0.000 | -1.480 | 0.146 | -2.939 | 0.005 | 1.047 | 0.295 | -7.188 | 0.000 |
| X ₁₄ | 净资产收益率 | 1.188 | 0.235 | 3.704 | 0.000 | 2.224 | 0.026 | -7.261 | 0.000 | -4.272 | 0.000 |
| X ₁₅ | 营业净利率 | 0.296 | 0.767 | -3.381 | 0.001 | -1.093 | 0.275 | 5.614 | 0.000 | -4.996 | 0.000 |
| X ₁₆ | 净利润现金净含量 | 0.622 | 0.537 | -1.353 | 0.183 | 1.717 | 0.086 | -1.976 | 0.048 | -1.477 | 0.147 |
| X ₁₇ | 销售收入现金含量 | 1.182 | 0.244 | 5.555 | 0.000 | -1.280 | 0.201 | -1.624 | 0.105 | -2.495 | 0.013 |
| X ₁₈ | 全部现金回收率 | -1.029 | 0.310 | -1.927 | 0.061 | -0.378 | 0.708 | -6.578 | 0.000 | -4.451 | 0.000 |
| X ₁₉ | 营运指数 | 1.203 | 0.234 | -2.870 | 0.004 | 2.205 | 0.028 | -1.444 | 0.149 | -0.034 | 0.973 |
| X ₂₀ | 财务杠杆 | -0.530 | 0.597 | -0.650 | 0.516 | -0.855 | 0.394 | 3.364 | 0.001 | -0.596 | 0.554 |
| X ₂₁ | 经营杠杆 | 1.265 | 0.210 | 0.139 | 0.890 | -2.517 | 0.015 | -4.394 | 0.000 | -0.383 | 0.703 |
| X ₂₂ | 综合杠杆 | -0.449 | 0.655 | -0.703 | 0.482 | -1.488 | 0.139 | 4.328 | 0.000 | -0.407 | 0.686 |
| X ₂₃ | 总资产增长率 | -5.064 | 0.000 | -3.486 | 0.001 | -2.966 | 0.003 | -1.942 | 0.058 | -6.741 | 0.000 |
| X ₂₄ | 净利润增长率 | 0.401 | 0.689 | -0.842 | 0.400 | 0.750 | 0.453 | -1.086 | 0.283 | -2.361 | 0.018 |
| X ₂₅ | 可持续增长率 | -3.346 | 0.001 | -0.405 | 0.686 | -1.300 | 0.198 | 2.397 | 0.017 | -6.319 | 0.000 |
| X ₂₆ | 所有者权益增长率 | -2.545 | 0.015 | -3.761 | 0.000 | -4.230 | 0.000 | 0.823 | 0.411 | -4.603 | 0.000 |
| X ₂₇ | 每股净资产增长率 | -1.623 | 0.105 | 0.204 | 0.839 | 0.264 | 0.792 | 0.395 | 0.695 | 0.396 | 0.692 |

当 P 值(Sig.)显示为 0.000 或小于 0.05 时, 检验结果有统计学意义, 统计结论为按 $\alpha=0.05$ 的标准拒绝原假设, 即被 ST 公司与正常公司在该财务指标上存在显著性差异。

(1) 总体来看, 被 ST 公司与正常公司在 T-6 年至 T-2 年分别有 11、12、13、16、15 个财务指标有显著差异。可见, 随着距离上市公司被特别处理的时间临近, 具有显著性差异的预测指标数量呈上升趋势。有四项指标在五个预测年度均没有通过显著性检验, 表明这四项预测指标在两类公司之间始终没有差异, 分别为: X₄(利息保障倍数)、X₇(应收账款周转率)、X₉(流动资产周转率)、X₂₇(每股净资产增长率)。因此, 将这四项指标从预测指标体系中剔除。

(2) 在偿债能力方面, ST 公司短期和长期偿债能力明显不如非 ST 公司, 尤其是反映短期偿债能力的流动比率和速动比率, 连续五年低于非 ST 公司。在经营能力方面, 存货周转率在发生财务危机的早期并没有表现出差异, 而是在公司被 ST 处理的前两年出现异常。在盈利能力方面, ST 公司的盈利能力相对较差。综合性比较强的资产报酬率和净资产收益率两个指标表现尤为明显。在现金流方面, 除距离被 ST 处理时间最长的 T-6 年(2009 年), 其余四年 ST 公司现金流动性均低于非 ST 公司。在风险水平方面, 除 T-3 年(2012 年), 其余各年反映风险水平的财务指标均没有明显差异。因此, 将风险水平纳入预警指标体系, 可能并不具有理论上的可行

□ 改革与发展

性。在发展能力方面,除净利润增长率外,其余预测指标均在多个年度存在显著差异,说明发展能力不足是上市公司被特别处理的重要影响因素。

基于以上分析,本文剔除了五年中均没有显著差异的指标和反映风险水平的指标,最终保留 X_1 、 X_2 、 X_3 、 X_5 、 X_6 、 X_8 、 X_{10} 、 X_{11} 、 X_{12} 、 X_{13} 、 X_{14} 、 X_{15} 、 X_{17} 、 X_{18} 、 X_{19} 、 X_{23} 、 X_{25} 、 X_{26} 共计 18 个预测指标作为进一步分析的变量。

2. 因子分析提取模型自变量。经过显著性检验,本文保留了 18 个预测指标。但是这些指标间存在较强的相关关系,如果直接作为自变量纳入模型,不但模型复杂,还会因为变量间存在的多重共线性而引起较大的误差。因此考虑利用因子分析法提取变量信息,减少分析维度,使问题更加简单直观。

(1) 确定变量结构是否适合应用因子分析法。本文采用 KMO 检验和 Bartlett 球形检验来判断变量结构是否适合采用因子分析法。各年度检验结果如表 3 所示:

表 3 KMO 检验和 Bartlett 球形检验结果

| 指标名称 | T-6 (2009年) | T-5 (2010年) | T-4 (2011年) | T-3 (2012年) | T-2 (2013年) |
|---------------|----------------|----------------|----------------|----------------|----------------|
| KMO 统计量 | 0.583 | 0.544 | 0.561 | 0.535 | 0.646 |
| Bartlett 球形检验 | 15910.866 | 27549.370 | 28888.017 | 22365.231 | 30340.999 |
| Sig. | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

由 Bartlett 球形检验可以看出, P 值(Sig.)均小于 0.05,说明变量间具有较强的相关性。但是 KMO 统计量均小于 0.7,说明各变量间信息重叠程度可能不是特别高,仍然值得尝试。综合以上两种检验结果,可认为本文的变量结构适合运用因子分析法。

(2) 确定公因子。以样本公司 T-2 年(2013 年)的数据为例,运用主成分分析法,依据特征值大于 1 的原则,提取了 6 个公因子,6 个主成分的累计方差贡献率达到 67.11%,基本包含了全部测量指标所具有的主要信息。具体结果如表 4 所示:

表 4 主成分及贡献率

| 主成分 | 特征值 | 贡献率(%) | 累积贡献率(%) |
|-----|-------|--------|----------|
| 1 | 3.890 | 21.614 | 21.614 |
| 2 | 3.174 | 17.633 | 39.246 |
| 3 | 1.657 | 9.203 | 48.449 |
| 4 | 1.211 | 6.726 | 55.176 |
| 5 | 1.088 | 6.042 | 61.218 |
| 6 | 1.060 | 5.888 | 67.106 |

提取公共因子的同时,可得到初始因子载荷矩阵。但初始因子模型的载荷矩阵往往难以对公因子做出合理的解释。此时,可以对初始因子载荷矩阵进行旋转,旋转后的因子载荷矩阵如表 5 所示。

表 5 旋转后的因子载荷矩阵

| 指标名称 | 公因子 | | | | | |
|----------|----------------|----------------|----------------|----------------|----------------|----------------|
| | F ₁ | F ₂ | F ₃ | F ₄ | F ₅ | F ₆ |
| 营业净利率 | 0.908 | | | | | |
| 全部现金回收率 | -0.868 | | | | | |
| 资产报酬率 | 0.867 | | | | | |
| 总资产净利润率 | 0.866 | | | | | |
| 固定资产周转率 | 0.861 | | | | | |
| 速动比率 | | 0.989 | | | | |
| 流动比率 | | 0.986 | | | | |
| 现金比率 | | 0.960 | | | | |
| 资产负债率 | | -0.441 | | 0.440 | 0.319 | |
| 所有者权益增长率 | | | 0.887 | | | |
| 总资产增长率 | | | 0.874 | | | |
| 产权比率 | | | | 0.739 | | |
| 净资产收益率 | | | | -0.558 | 0.383 | |
| 总资产周转率 | | | | | 0.723 | |
| 营运指数 | | | | | | |
| 销售收入现金含量 | | | | | | 0.682 |
| 存货周转率 | | | | | | -0.634 |
| 可持续增长率 | | | | 0.342 | | 0.347 |

为了使结果更清晰易懂,表格按照系数大小进行了排序,并且过小的系数(绝对值小于 0.3)被抑制输出。由表 5 可以看出:公因子 F₁ 在指标营业净利率、资产报酬率、总资产净利润率上有较大的载荷,它们主要反映了上市公司的盈利能力,可以命名为盈利因子。同理,公因子 F₂ 至公因子 F₆ 可分别命名为短期偿债因子、发展因子、长期偿债因子、经营因子、现金流因子。提取结果表明,应用因子分析法所提取的公因子不仅包括原始的偿债、经营、盈利、现金流和发展能力五个方面,而且将偿债能力进一步区分为短期偿债能力和长期偿债能力。这样既降低了分析维度,又使得提取的信息含义更加清晰。

基于上述分析结果,可得到用来代替 27 个原始变量的 6 个新变量,即最终取 F₁(盈利因子)、F₂(短期偿债因子)、F₃(发展因子)、F₄(长期偿债因子)、F₅(经营因子)、F₆(现金流因子)作为财务预警模型的自变量。

(二) 预警模型分析

本文以样本公司是否为 ST 公司为因变量,以因子分析模型提取的 6 个公因子为自变量,以公司被特别处理前五年的数据为基础,利用 Logistic 回归、决策树和 BP 神经网络分别建模如下:

1. 二分类 Logistic 回归模型。

(1) 模型拟合优度分析。Logistic 模型可通过 -2 倍的对数似然值来表示模型的拟合效果,其值越小,说明模型的拟合效果越好。各个预测年度模型的拟合效果如表 6 所示。

表 6 模型拟合优度检验

| 年份 | -2 对数似然值 | Cox & Snell R ² | Nagelkerke R ² |
|------------|----------------------|----------------------------|---------------------------|
| T-6(2009年) | 306.403 ^a | 0.045 | 0.159 |
| T-5(2010年) | 318.053 ^a | 0.038 | 0.158 |
| T-4(2011年) | 306.996 ^a | 0.040 | 0.175 |
| T-3(2012年) | 291.385 ^a | 0.047 | 0.211 |
| T-2(2013年) | 275.697 | 0.054 | 0.247 |

从模型的拟合优度来看,在T-5年至T-2年,-2 对数似然值越来越小,说明随着距离公司被ST的时间越短,Logistic模型的拟合效果越好。

Cox & Snell R²和Nagelkerke R²从不同的角度反映了当前模型中自变量解释的因变量变异占因变量总变异的比列,类似线性回归模型中的决定系数,从表6中可以看出,Cox & Snell R²和Nagelkerke R²值 T-2>T-3>T-4>T-5,说明在T-2年,自变量对因变量的解释能力最强。

(2)模型检验结果。模型检验结果如表7所示。

2. 基于C&RT算法的决策树模型。

表 7 Logistic 回归模型检验结果

| 年份 | 实际类型 | 预测类型 | | 判别正确率 | 总正确率 |
|----------------|------|------|------|-------|-------|
| | | ST公司 | 正常公司 | | |
| T-6 (2009年) | ST公司 | 24 | 13 | 62.9% | 70.1% |
| | 正常公司 | 272 | 650 | 70.5% | |
| T-5 (2010年) | ST公司 | 23 | 16 | 60.2% | 74.4% |
| | 正常公司 | 207 | 715 | 77.6% | |
| T-4 (2011年) | ST公司 | 25 | 14 | 63.2% | 75.0% |
| | 正常公司 | 225 | 697 | 75.6% | |
| T-3 (2012年) | ST公司 | 20 | 19 | 50.8% | 76.6% |
| | 正常公司 | 205 | 717 | 77.8% | |
| T-2 (2013年) | ST公司 | 27 | 12 | 70.0% | 87.2% |
| | 正常公司 | 84 | 838 | 90.9% | |

(1)模型的建立。C&RT是目前应用最广泛的决策树模型算法,适用于任意类型的变量。本文首先采用C&RT算法将样本公司分成多个亚群。以T-2年的数据为例,生成的决策树模型如图1所示:

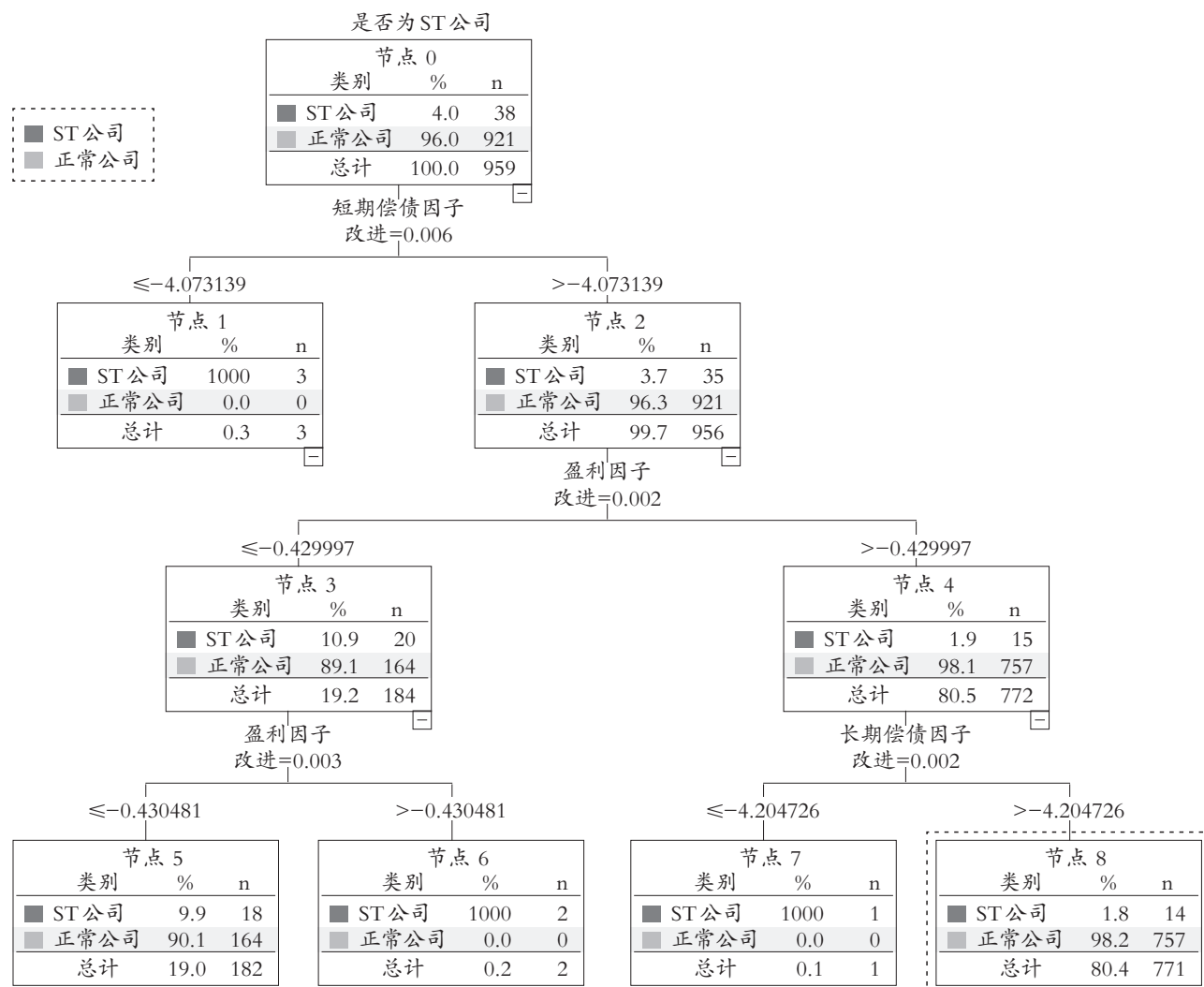


图 1 决策树模型分析结果

□ 改革与发展

可见,根节点中ST公司所占比例为4%。模型的第一层是按照短期偿债因子进行拆分,拆分点为是否超过-4.07,短期偿债因子低于-4.07的上市公司,全部为ST公司,超过-4.07的余下公司中,ST公司占比下降为3.7%。以此类推,不断重复分枝,获得越来越纯的节点样本。利用决策树模型进行预测时,可归纳出希望预测的目标类别的特征属性。例如,节点1、节点6、节点7中ST公司占比均为100%,那么该节点即代表了ST类公司,沿着树的路径走回根节点,沿路所经过的节点的属性即为大多数ST公司所共有的属性。

(2)模型的预测结果。决策树模型的预测结果如表8所示:

表8 决策树模型预测结果

| 年份 | 实际类型 | 预测类型 | | 判别正确率 | 总正确率 |
|----------------|------|------|------|-------|-------|
| | | ST公司 | 正常公司 | | |
| T-6 (2009年) | ST公司 | 26 | 13 | 67.6% | 76.3% |
| | 正常公司 | 215 | 707 | 76.7% | |
| T-5 (2010年) | ST公司 | 27 | 12 | 70.3% | 77.6% |
| | 正常公司 | 176 | 746 | 80.9% | |
| T-4 (2011年) | ST公司 | 28 | 11 | 72.1% | 80.0% |
| | 正常公司 | 181 | 741 | 80.4% | |
| T-3 (2012年) | ST公司 | 24 | 15 | 62.5% | 74.4% |
| | 正常公司 | 230 | 692 | 75.0% | |
| T-2 (2013年) | ST公司 | 27 | 12 | 69.6% | 86.0% |
| | 正常公司 | 122 | 800 | 86.8% | |

3. BP神经网络模型。在结构上,神经网络包括输入层、输出层和隐含层;输入层的每个节点对应各预测变量,相当于统计学中的自变量;输出层节点对应目标变量,其信号强弱相当于因变量预测结果;在输入层和输出层之间是隐含层,相当于统计学中采用的统计模型(对神经网络使用者不可见)。信息由输入层向输出层不断传递,中间神经元根据某种不可知的准则不断调整对接收信号的处理方式,最终使输出层得到想要的信息。以T-2年(2013年)为例,图2为所建立的BP神经网络模型:

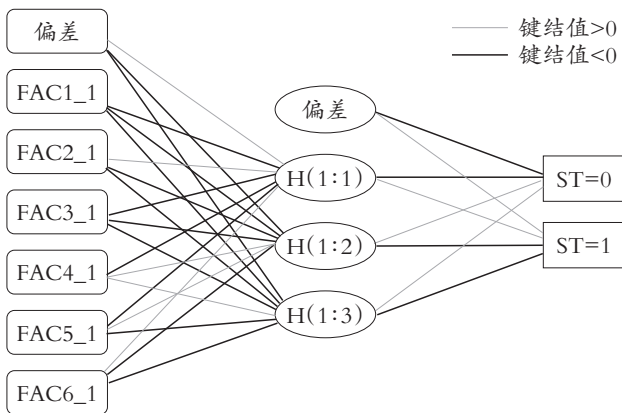


图2 BP神经网络模型结构

该模型输入层纳入了6个神经元,即因子分析提取的6个公因子,隐含层中包括了三个隐单元,相应的两分类因变量以两个节点的形式构成了输出层。图中相邻两个节点两两链接,连接线的粗细代表权重绝对值的大小,从图中可以看出,自变量 F_1 、 F_3 对模型的贡献明显比较大,并且通过隐含层的H(1:1)节点和输出层的ST=0有较强的链接权重,这提示自变量 F_1 、 F_3 对应的盈利因子和发展因子在两类企业中区分度较大。

为防止过度拟合,本文按照7:3的比例将总样本拆分成训练集和验证集,并重点关注验证集的预测结果。样本的预测分类结果如表9所示:

表9 BP神经网络模型预测结果

| | T-6 (2009年) | T-5 (2010年) | T-4 (2011年) | T-3 (2012年) | T-2 (2013年) |
|-----|----------------|----------------|----------------|----------------|----------------|
| 训练集 | 75% | 77.2% | 81.2% | 77.4% | 87.1% |
| 测试集 | 85% | 76.5% | 85% | 77.6% | 86.5% |

4. 模型评价。

(1)模型比较。由上述分析可知,Logistic回归、决策树和BP神经网络在实际应用中各有优缺点,具体如表10所示:

表10 三种预测模型的优缺点

| | 传统统计方法 | 数据挖掘技术 | |
|----|---|---|---|
| | Logistic回归 | 决策树 | BP神经网络 |
| 优点 | 满足对分类数据的建模需求;可自动对自变量进行筛选 | 适用条件限制较少,应用范围广,适合于对各种复杂的联系进行分析;分析原理与所得结果简单易懂;可以有效地处理数据中自变量缺失的问题 | 具有一定的容错性和很强的不确定性信息处理能力;结构特点和信息存储的分布式特点,使得神经网络具有较强的稳健性 |
| 缺点 | 模型的建立需满足假设条件,对样本量也有严格的要求;对错误资料的输入不具有容错性 | 不能对影响因素的作用大小进行精确的定量描述;对于结果的解释和应用过于灵活,没有严格的标准可循 | 可解释性差,只能作为一种“黑箱”方法加以应用,无法深入分析、探讨自变量和因变量的关联方式;无法区分有效信息和噪声,容易导致训练过度问题 |

从预测模型固有的优缺点来看,具体选择哪个模型要根据研究目的而定。如果研究目的是深入探讨企业发生财务危机的原因,可选择Logistic回归模型,自动筛选出对企业陷入财务困境有影响作用的自变量;如果研究目的是探讨发生财务危机企业共有的特征,可以考虑决策树模型,从目标类别所在节点,沿着树的路径走回根节点,沿路所经过的节点的属性即为大多数ST公司所共有的属性;如果企业只是想预测是否会发生财务危机这个结果,而不关心中间的过程,可

以选择BP神经网络模型。

(2)模型预测结果对比分析。根据上述研究结果,可汇总三种模型预测的准确率,如表11所示:

表 11 三种模型预测准确率统计

| 模型 | 预测准确率 | | | | |
|-------------|----------------|----------------|----------------|----------------|----------------|
| | T-6 (2009年) | T-5 (2010年) | T-4 (2011年) | T-3 (2012年) | T-2 (2013年) |
| Logistic 回归 | 70.1% | 74.4% | 75% | 76.6% | 87.2% |
| 决策树 | 76.3% | 77.6% | 80% | 74.4% | 86% |
| BP神经网络 | 85% | 76.5% | 85% | 77.6% | 86.5% |

整体来看,采用 Logistic 回归、决策树和BP神经网络构建的预警模型对ST发生前2~6年的数据进行财务预警具有一定的正确性,并且随着ST发生时间的临近,预警准确率呈上升趋势。

从各年预测准确率来看,基于数据挖掘技术的决策树和BP神经网络模型准确率要优于传统统计分析的 Logistic 回归模型,尤其是在预测的早期。在T-6年,BP神经网络模型的准确率最高,为85%;其次是决策树模型,为76.3%;Logistic 回归模型准确率最低,仅为70.1%。随着ST发生时间的临近,决策树和BP神经网络模型在准确率上的优势逐渐消退,在T-2年,甚至低于 Logistic 回归模型的准确率。这说明BP神经网络和决策树方法适合应用于长期预警,Logistic 回归则更适合应用于短期预警。

五、结论

本文以制造业上市公司为研究对象,利用 Logistic 回归、决策树和 BP神经网络分别构建了财务预警模型,根据预测结果得出以下研究结论:

从适用范围来讲,Logistic 回归模型适用于深入探讨企业发生财务危机的原因;决策树模型适用于研究发生财务危机企业共有的特征;BP神经网络模型适用于直接判断企业是否会发生财务危机。

从预测能力来看,决策树和 BP神经网络模型早期的预测能力优于 Logistic 回归模型,随着ST发生时间的临近,这种优势逐渐弱化。短期内,三种模型的预测能力不相上下,在企业被特别处理的前两年,三者准确率均在85%以上。

综上所述,Logistic 回归、决策树和BP神经网络在判断制

造业上市公司的财务状况方面,均具有一定的应用价值。企业实际建立财务预警系统时,可根据预警模型的特点和本企业的预警偏好,选择合适的模型。

主要参考文献:

Altman E. I.. Financial Ratios Discriminant Analysis and the Prediction of Corporate Bankruptcy[J]. The Journal of Finance,1968(9).

Tzong-Huei Lin. A Cross Model Study of Corporate Financial Distress Prediction in Taiwan: Multiple Discriminant Analysis, Logit, Probit and Neural Networks Models [J]. Neurocomputing,2009(16).

Jia-Jane Shuai, Han-Lin Li. Using Rough Set and Worst Practice DEA in Business Failure Prediction[J]. Lecture Notes in Computer Science,2005(10).

吴星泽. 财务预警的非财务观[J]. 当代财经,2010(4).

Ohlson J. A.. Financial Ratios and the Probabilistic Prediction of Bankruptcy [J]. Journal of Accounting Research, 1980(19).

陈静. 上市公司财务恶化预测的实证分析[J]. 会计研究,1999(4).

吴世农,卢贤义. 我国上市公司财务风险的预测模型研究[J]. 经济研究,2001(6).

Wilson R. L., Sharda R.. Bankruptcy Prediction Using Neural Networks [J]. Decision Support System,1994(11).

马若微. 基于粗糙集与信息熵的上市公司财务困境预警指标的确定[J]. 当代经济科学,2005(2).

姚靠华,蒋艳辉. 基于决策树的财务预警[J]. 系统工程,2005(10).

李艳玲,宋喜莲. 基于数据挖掘的上市公司财务危机预警研究[J]. 沈阳师范大学学报(自然科学版),2014(2).

张亮,张玲玲,陈懿冰,腾伟丽. 基于信息融合的数据挖掘方法在公司财务预警中的应用[J]. 中国管理科学,2015(10).

王宗胜,尚娇娇. 我国制造业上市公司财务困境预警分析[J]. 统计与决策,2015(3).

作者单位:江苏师范大学商学院,江苏徐州221116