

文章编号:1003-207(2015)10-0170-07

DOI:10.16381/j.cnki.issn1003-207x.2015.10.020

基于信息融合的数据挖掘方法在公司财务预警中的应用

张亮^{1,2}, 张玲玲^{1,2,3}, 陈懿冰^{2,4}, 滕伟丽⁵

(1. 中国科学院大学管理学院, 北京 100190;

2. 中国科学院虚拟经济与数据科学研究中心, 北京 100190;

3. 中国科学院大数据挖掘与知识管理重点实验室, 北京 100190;

4. 中国科学院科技政策与管理科学研究所, 北京 100190;

5. 英国诺丁汉特伦特大学诺丁汉商学院, 诺丁汉伯顿街 NG14BU)

摘要:目前越来越多的数据挖掘方法被用于风险预警中,决策树、支持向量机、神经网络、Logistic 回归等方法在风险预警中都表现出了较好的特性和预警效果,但是不同数据挖掘分类方法得到的结果不同,往往导致预警结果的不一致,因此也会存在一定风险。本文引入信息融合技术对不同数据挖掘分类方法得到的结果进行融合处理得到最优的结果,解决了不同数据挖掘方法得到的结果不一致问题。文章在 SVM 和 Logistic 回归的数据挖掘模型基础上建立基于信息融合的公司财务预警模型,提高了财务预警准确率,并且保留了原数据挖掘方法在分类预测上的优势。在实证研究中,论文选取了中国制造业的上市公司作为研究对象,在 SVM 和 Logistic 回归两种数据挖掘模型的基础上利用信息融合方法建立了财务预警模型,实证结果表明,基于信息融合的数据挖掘方法的预测准确率要高于单独的 SVM 和 Logistic 回归两种方法。

关键词:财务预警;信息融合;支持向量机;Logistic 回归

中图分类号:F224

文献标识码:A

1 引言

公司财务预警是指财务失败预警,针对公司经营中的经营计划、财务报表以及其他生产经营中的会计数据,利用统计学、金融学以及财会学理论中的比率分析因素分析等方法,分析公司经营的财务状况,发现公司在经营过程中潜在的财务风险,并且期望在风险发生之前做出警示,提前做出决策控制或避免风险的发生。近年来,上市公司财务预警成为研究的焦点,数据挖掘的方法也逐渐成为分析上市公司财务状况的重要手段。

国内外学者运用了不同的方法对财务预警进行了比较研究。在早期 Altman^[1]等在 Z-score 模型的基础上建立了改进的 ZETA 模型用于财务预警,并且提高了预测准确率。Ohlson^[2]将统计中 Logistic

模型引入财务危机的预测中,其结果比前两个模型的预测准确率都得到了提高。Odom^[3]等最早使用神经网络技术进行财务危机预测,他们选取 5 项财务比率指标建立了神经网络预警模型,研究发现使用神经网络的方法对公司财务危机的预测率要优于基于统计的方法。杨淑娥^[4]等引入了面板数据构建 BP 神经网络模型预测了上市公司的财务状况。研究表明面板数据对 BP 神经网络预警的分析显示出稳定、连续的预测性能,并且预测能力有较大的提高。Gestel^[6]选取荷兰破产企业为研究对象,采用支持向量机(Support Vector Machine, SVM)进行财务状况预测,通过贝叶斯推理得到核参数,预测结果为 88.39%,比以往基于统计的预警模型有更高的预测精度。Shin^[7]重点研究了 SVM 和 BP 神经网络的性能差异,实验表明在小样本情况下,BP 神经网络在测试样本上预测效果下降,而 SVM 预测效果会更好。孙洁,李辉等^[14]提出了基于滚动时间窗口支持向量机的财务困境预测动态建模新方法。通过模拟时间推移过程,对我国上市公司展开实证

收稿日期:2014-04-29; 修订日期:2015-02-02

基金项目:国家自然科学基金资助项目(71071151,71471169)

作者简介:张亮(1975—),男(汉族),安徽人,中国科学院大学管理学院,硕士,研究方向:数据挖掘、金融管理。

研究,结果对未来企业财务困境的预测效果明显优于静态 SVM 模型。韩建光等^[15]在支持向量机、多元判别分析、Logistic 回归、分类和回归树等多种分类学习算法的基础上,提出了基于精度前向搜索和后剪枝的多特征子集组合分类器财务困境预测方法,该方法构建的组合系统的分类预测精度明显高于个体最优模型。刘遵雄和郑淑娟等^[16]将 L_1 范数正则化 Logistic 回归模型用于上市公司财务危机预报,实验结果表明 L_1 正则化 Logistic 回归模型的有效性,其在保证模型预测精度的同时提高模型的解释性。周宗放,肖珉^[17]运用了 Logistic 模型、支持向量机 SVM 和 CLL 模型三种方法,验证了引入企业集团的特征类指标和治理类指标能显著提高企业集团信用风险评估的准确性。孟杰^[18]利用随机森林方法建立了我国上市公司的财务失败预警模型,并与 Logistic 回归、支持向量机、CART 分类树以及神经网络方法进行对比,表明随机森林模型预测精度更高、稳健性更好。辛金国,关建清^[19]运用神经网络模型与决策树模型分别构建了民营上市公司绩效评价体系,实证分析表明神经网络模型对民营上市公司绩效评价研究非常有效,并且神经网络模型在民营上市公司绩效评价研究上要优于决策树模型。杨海珍,荆中博^[20]在用单变量检验方法检验备选财务指标有效性的基础上,建立了加权 Logit 模型,利用模型贡献度指标分析财务因素的显著性和影响程度。文柯^[21]采用因子分析和 Logistic 回归构建我国企业产融结合风险预警模型,研究结果表明,产业扩张速度、财务费用占比与产融结合后陷入财务困境正相关;企业偿债能力越强、盈利水平越高、营运能力越强,产融结合后陷入财务困境的概率越低。

综上所述,现有的文献研究大多都是通过选取企业经营的财务和会计数据作为模型的输入指标,不断改进和利用新模型提高财务危机的预测准确性。当前决策树、贝叶斯理论、神经网络以及现今流行的支持向量机等数据挖掘方法在财务危机的预测中都获得了较好的效果。但不足之处是由于不同数据挖掘方法的特点不同,各种方法在适用条件上的差异使得不同的方法得到的结果不同,经常会存在结果的不一致性,具有一定的局限性,最终的挖掘结果支持决策性不高。因此,对数据挖掘结果进行“二次”处理以提高最终挖掘结果的预期效果和提高决策支持性更加重要。针对财务危机预测研究,本文主要贡献是在传统数据挖掘方法的基础上建立财务预警模型,并引入信息融合技术,采用 DS 证据理论

算法对 Logistic 回归和支持向量机建立的财务预警模型得到的不同结果进行再次融合处理,通过对我国上市公司的实证研究对比,表明该方法对处理数据挖掘得到不同的预测结果具有一定有效性,并且相对单一的数据挖掘方法提高了最终预测准确率。

2 基于信息融合的财务预警模型

目前针对企业财务预警的预测研究,很多学者通过改进现有的经典模型方法或运用新的模型方法提高预测结果的准确率,但是从单个模型方法出发对财务预警的研究使得不同模型方法得到的预测结果也会不同,财务数据特性的不同算也使得模型方法的运用具有一定的局限性,预测的准确率也受到影响。因此,为了较好的融合不同算法的优势,建立综合的评估预测模型,提高决策支持性,本研究在数据挖掘的基础上,针对不同数据挖掘结果运用信息融合技术,融合不同算法的结果提高预测结果的可靠性。本研究选取分类模型中分类效果较好的支持向量机模型和 Logistic 回归模型作为数据挖掘的方法,然后运用信息融合技术对两种数据挖掘的结果进行融合计算,优化最终的决策结果。

本文研究上市公司的财务预警问题,目前沪深证券交易所根据股票上市的规则,针对上市公司的财务状况或其他情况出现异常时,对其股票交易特别处理。被特别处理的上市公司股票名称前加上“ST”标识,其股票的当日价格涨跌幅均限制为 5%,并且中期报告必须审计。本文以上市公司中被“ST”公司和非“ST”公司作为研究对象,分析上市公司的财务数据,利用数据挖掘技术建立财务预警模型,预测公司未来的财务状况。

2.1 基于 SVM 的财务预警模型

支持向量机(Support Vector Machine)是由 Vapnik 提出的一种可以很好的解决线性和非线性分类问题的数据挖掘方法。SVM 的主要思想是通过使用非线性映射将低维输入空间线性不可分的样本转化为高维特征空间使其线性可分,在高维的特征空间里寻找线性最优分割超平面。

在支持向量机模型中,数据集 D 为 $(x_1, y_1), (x_2, y_2), \dots, (x_{|N|}, y_{|N|})$, 其中 x_i 为数据集合的属性, y_i 为每一条数据项的识别标签,支持向量机通过在平面上寻找两类数据集间隔最大的分界线作为划分数据集的分界线,其基本原理如图 1 所示。

SVM 的主要算法思想如下,如果在平面中训练数据集是线性可分的,则决策函数为:

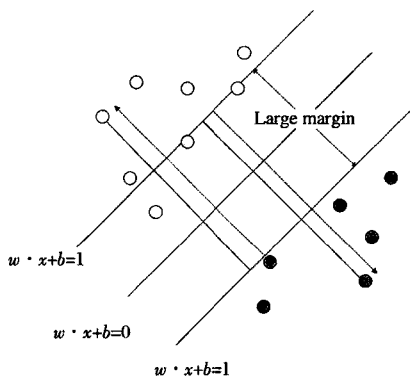


图1 SVM的超平面和架构

$$y_i = w \cdot x + b \quad (1.1)$$

其中 w 是权重向量, $w = (w_1, w_2, \dots, w_N)$, b 是偏移向量。如果 $y_i > 0$, 则数据集被分为 +1 类, 如果 $y_i < 0$, 则数据集被分为 -1 类。如果数据集是非线性可分的, 则决策函数为:

$$\text{Min} \left(\frac{1}{2} w^T w + C \sum_{i=1}^N \delta_i \right) \quad (2)$$

$$y_i(w \cdot x + b) \geq 1 - \delta_i, \delta_i \geq 0, i = 1, 2, \dots, N$$

其中, C 是惩罚因子决定了正负类别的最大间距和最小分类误差, δ_i 是非负松弛变量。

在财务预警的 SVM 模型研究中, 公司的财务数据指标作为模型的输入向量, 也是模型分类的属性集, 通过 SVM 的机器学习算法, 在空间中寻找一个超平面可以将公司进行分类。以“ST”作为公司分类的标签, 建立 SVM 分类模型对上市公司进行财务危机预测分类。在 SVM 模型得出的结果形式为 (y_1, y_2) , 例如 $(1, 0)$ 和 $(0, 1)$, 当 $y_1 = 1$ 表示该公司被 SVM 模型预测为将来可能会成为“ST”公司, $y_2 = 1$ 则表示该预测公司将来财务状况保持良好状况。

2.2 基于 Logistic 的财务预警模型

Logistic 回归是一个可以解决分类问题的高效的数据挖掘方法, 并且可以用于预测分析。Logistic 模型通过极大似然估计的方法建立一个线性回归分类模型, 可以对二值变量或多值变量分类的数据集进行分类。

在二值变量的 Logistic 回归模型中, y_i 服从伯努利概率分布函数, 是一个二值的回归输出结果, 分别为 $y = +1$ 和 $y = -1$, 并且 Logistic 回归模型给出了 y 取值的概率大小。在 Logistic 回归的函数中, x_i 为解释自变量, y_i 服从以下概率分布:

$$P(y_i = 1 | x_i) = P[\varepsilon_i \leq (\alpha + \beta x_i)] = \frac{1}{1 + e^{\alpha + \beta x_i}} \quad (3)$$

其中参数 $\beta = (\beta_0, \beta_1)$ 是一个 $n \times 1$ 的向量, β_0 是一个常量, β_1 对应不同解释自变量的系数向量。Logistic 函数分布如图 2 所示。

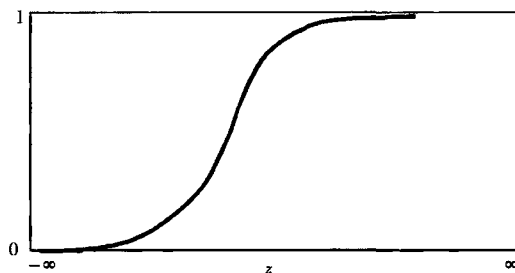


图2 Logistic 函数分布图

在财务预警的 Logistic 模型中, 公司财务指标作为 Logistic 回归的解释变量 X , 以“ST”作为分类标签, 被定义为: $y = 1$ 为“ST”公司, 否则为非“ST”公司。在 Logistic 模型中, 模型是根据对解释变量回归分析, 得出公司被预测为“ST”结果概率发生比, 然后根据发生概率的大小, 判断公司是否会成为“ST”公司。Logistic 回归模型建立的财务危机预测模型为:

$$y_i = \ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (1.4)$$

经过计算转换为:

$$p_i(y_i = 1 | X) = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)]} \quad (1.5)$$

Logistic 模型中给出了财务预警的两个结果, 一个是判断预测公司是否会成为“ST”公司, 由 y 表示, 一个是对此判断结果的概率支持, 由 p 表示, Logistic 模型对分类预测问题给出了良好的解释。

2.3 基于信息融合的财务预警模型

信息融合技术是将不同信息源的信息通过融合算法解决单一结果的决策问题, 群体决策等。目前信息融合技术具有很多算法, 例如加权平均法、多贝叶斯估计法、证据推理方法和模糊逻辑等。其中, DS 证据理论是信息融合最经典的算法, DS 证据理论通过证据融合规则将不同的信息源信息融合计算得到最后的单一结果。DS 证据理论已经成为信息融合领域一种重要的研究方法。

DS 证据理论一种基于不确定性和不完整信息的数学推理算法, 他来自不同信息源的证据进行合成处理, 得到最后的信任度。在信息融合的计算

过程中,信息融合的整体判断框架被称为识别框架,融合的基础是基本概率分布,然后根据融合规则进行信息融合计算,信任函数是描述融合结果对某个假设的支持程度,即信任程度。

首先,假设 U 为一些互斥且穷举的元素组成的命题的非空集合,被称为辨识框架, 2^U 是由 U 的所有子集构成的幂集合,表示对某个事件判断的所有可能的假设命题。例如,识别框架 $U = \{U_1, U_2, U_3, \dots, U_n\}$ 中有 N 个假设,每一个假设表示其出现的状态。DS 证据融合过程是基于每一个假设的信任度计算最终的信任度。基本概率分配函数(basic probability assignment, BPA)又被称为信任结构,有时也被成为 mass 函数,简称为 m 函数。 m 函数是从 2^U 到 $[0, 1]$ 上的映射,表示幂集合里每个元素的基本信任测度,也分别为对其假设的支持度。该基本概率分布有下列性质:

$$(1) m(\emptyset) = 0$$

$$(2) \sum_{A \subseteq U} m(A) = 1$$

对于集合 A 为 U 集合中的子集,则其基本概率分配函数为 $m(A)$,表示命题假设 A 的基本信任水平,并且非空真子集 A 的基本信任水平在 $[0, 1]$ 之间,集合 A 被称为焦元。

信任函数或信度函数(Belief Function, Bel)是 DS 证据融合理论中重要的函数,它表示在识别框架 U 下,对任意一个假设的基本信任测度区间的下限,其函数表达式为:

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad (1.6)$$

$Bel(A)$ 表示对 A 的总的信任,是 U 上的信任函数。在证据融合计算的过程中,直接计算某假设的支持度,而不需计算包含该假设中的所有可能的支持度。

DS 证据融合理论的融合规则可以通过融合公式进行计算。在辨识框架 U 上有两个信任函数分别为 Bel_1 和 Bel_2 ,其信任程度分别为 m_1 和 m_2 ,信任函数中对应的焦元分别为 A_1, A_2, \dots, A_n 和 B_1, B_2, \dots, B_n ,融合的计算公式则为:

$$m(C) = \begin{cases} \frac{\sum_{A_i \cap B_j = C} m_1(A_i) * m_2(B_j)}{1 - \sum_{A_i \cap B_j = \emptyset} m_1(A_i) * m_2(B_j)}, & A \cap B \neq \emptyset \\ 0, & A \cap B = \emptyset \end{cases} \quad (1.7)$$

其中, $\sum_{A_i \cap B_j = \emptyset} m_1(A_i) * m_2(B_j) < 1, \forall C \subseteq U,$

$$C \neq \emptyset, m(\emptyset) = 0$$

$$\text{令 } K = \sum_{A_i \cap B_j = \emptyset} m_1(A_i) * m_2(B_j), K \text{ 表示各证}$$

据间的冲突程度,而 $1/(1-K)$ 为正则化因子。由 $m(C)$ 得到的信任函数为 m_1 和 m_2 正交和,有时可记为 $m_1 \oplus m_2$,并且只有当 $K < 1$ 时该函数才成立存在。DS 证据融合规则是将不同证据联合的法则,对于同一个识别框架内的不同证据信任函数,如果所有的证据都不完全冲突,那么 DS 证据融合规则可以有有效的计算出一个联合证据的信任函数。

在本研究中,对 SVM 结果和 Logistic 结果运用上述算法进行融合计算。由这两个模型得到的结果是被预测的上市公司分为“ST”公司和非“ST”公司,其模型输出的结果为一个两维的结构,即 $(1, 0)$ 和 $(0, 1)$ 。在证据融合理论中可将其定义为识别框架, $?U = \{ST, \text{非} ST\}$ 。ST 表示“预测某公司会出现财务危机”的假设,非 ST 则表示“预测某公司不会出现财务危机”的假设。在每个模型中的不仅会输出对公司财务状况预测的假设,而且会计算出由该模型对预测出结果的支持度,即信任度。SVM 和 Logistic 模型将每个公司进行分类预测,并且计算出分类预测的概率 $p = (p_0, p_1)$, p_0 表示公司被预测为财务正常的概率,即由该模型获得证据信任度对该公司被预测为非 ST 公司的支持度。 p_1 表示公司被预测为财务危机的概率,即由该模型获得证据信任度对该公司被预测为 ST 公司的支持度。由 SVM 模型得到的分类预测概率记为 $P_{SVM} = (p_0, p_1)$,其表示分别对假设 $H_1 = \{F_1 = \text{“ST 公司”}, F_2 = \text{“非 ST 公司”}\}$ 的基本信任度,信任函数为 $m_1(F_1)$ 和 $m_1(F_2)$ 。同样由 Logistic 模型得到的分类预测概率记为 $P_{logistic} = (p'_0, p'_1)$,其表示分别对假设 $H_1 = \{F_1 = \text{“ST 公司”}, F_2 = \text{“非 ST 公司”}\}$ 的基本信任度,信任函数为 $m_2(F_1)$ 和 $m_2(F_2)$ 。然后利用 DS 证据融合公式计算联合证据信任度,计算公式为:

$$m(C) = \frac{\sum_{F_i \cap F_j = C} m_1(F_i) * m_2(F_j)}{1 - \sum_{F_i \cap F_j = \emptyset} m_1(F_i) * m_2(F_j)}, (i, j = 1, 2) \quad (1.8)$$

其中, $\sum_{F_i \cap F_j = \emptyset} m_1(F_i) * m_2(F_j) < 1, \forall C \subseteq U,$
 $C \neq \emptyset, m(\emptyset) = 0, m_1(F_i) = p_i, m_2(F_j) = p'_j$
 $m(C)$ 是 DS 证据融合计算的信任程度,是基于 SVM 模型和 Logistic 模型得到的预测结果和其分

别模型的信任度得到的联合信任度,然后通过联合信任度预测公司未来的财务状况,即是否会成为 ST 公司。通过这样的融合过程,可以提高最终预测结果的可靠程度,这也是信息融合技术得出最后的决策结果。

3 实证分析

信息融合技术可以很好的融合数据挖掘技术的优势,并且改善数据挖掘的结果,提高决策可支持性。为了验证信息融合技术对数据挖掘结果改善的有效性,本部分进行了实证分析研究,选取我国制造业上市公司作为研究对象,运用 SVM 和 Logistic 数据挖掘模型对公司历史财务数据进行分析,预测公司未来发生财务危机的可能性。在前两种模型的基础上通过信息融合技术建立财务预警模型。实证分析中的数据来源于 WIND 数据库中所有上证和深证的制造业上市公司的公开财务数据,样本对象的选择原则为:如果该公司在当年 T 被“ST”,则取 T-2 年的财务数据作为分析数据。根据这样的原则最终选择了 92 家公司作为研究对象,其公开的财务数据作为模型的解释变量。

3.1 变量的选择

在现有的财务预警研究中,选取关键的财务指标数据对模型最终的预测结果具有重大的影响,很

多学者在选取财务预警模型的输入指标时均采用了公司经营中的财务和会计数据,表 1 列举了已有研究中财务预警的指标。

由于本文数据来源于上市公司的公开财务数据,包括偿债能力、每股指标、收益质量、盈利能力、运营能力和资本结构。鉴于对已有研究中关键指标的参考,并从统计的角度删除了数据缺失严重的指标,最后保留 22 个财务指标作为模型的输入数据,如表 2 所示。

3.2 基于信息融合模型的预测

本文将研究的样本数据随机分为两类,一类为训练集,用于财务预警模型的建立,一类为测试集,用于检验模型预测的准确率,训练集与测试集数据样本比例为 71:21。当模型的输出结果 $y_i = 1$ 时,则表明该公司被“ST”,公司可能面临财务危机的风险,当 $y_i = 0$ 时,则表明公司的财务状况正常,公司面临财务危机风险的可能性较小。将选取的 25 个模型变量分别运用 SVM 模型和 Logistic 模型建立财务预警模型,并且计算出各分类预测的概率。然后运用信息融合技术对 SVM 和 Logistic 模型预测的结果进行融合计算,最后在计算出预测的准确率,并与 SVM 和 Logistic 模型的准确率进行比较,各模型的预测结果以及准确率从比较如表 3 和表 4 所示。

表 1 财务预警指标^[2,3,5,8-13]

模型变量	使用的方法	作者
企业的财务结构、企业的规模、经营绩效、流动性	Logistic	Ohlson, 1980
营运资产/总资产、留存收益/总资产、息税前利润/总资产、权益市场价值/负债账面总价值、销售收入/总资产	神经网络	Odom, Sharda,
盈利能力、偿债能力、营运能力、成长能力和公司规模的 17 个财务指标	决策树	姚靠华, 蒋艳辉, 2005
流动比率、速动比率、库存流动资产、流动负债资产比、负债比率、流动负债比率、长期负债比率、权益资产比、流动资产周转率等 18 个指标	CHAID 模型	AliSerhan Koyuncugil, Nermin Ozgulbas, 2012
利润率比率、举债经营率、偿债能力比率、资产管理比率	SVM 模型	杨毓, 蒙肖莲, 2006
总资产净收益、有形资产增长率、股权净收益、净收入、现金流总债务比率、存货周转率、固定资产周转率等 42 个指标	SVM 模型	Jae H. Min, Young-Chan Lee, 2005
短期偿债能力、长期偿债能力、盈利能力、资产管理能力、主营业务鲜明程度、公司增长能力	BP 神经网络	杨淑娥, 黄礼, 2005
偿债能力、营运能力、成长能力、股东获利能力、盈利能力、现金能力	Logistic 模型	康晓玲, 张露, 2009
营运能力指标、流动能力指标、偿债能力指标、盈余管理指标、现金指标、信用能力	KMV 模型	谭久均, 2005

表 2 模型输入指标

财务指标	模型变量
偿债能力	流动比率、速动比率、股东权益/负债合计、息税前利润/负债合计、经营现金流净额/负债
每股指标	每股收益、每股收益(稀释)、每股经营现金流、每股净资产
盈利能力	销售毛利率、销售净利率、净资产收益率、总资产净利率、
运营能力	存货周转率、应收账款周转率、流动资产周转率、固定资产周转率、总资产周转率
资本结构	资产负债率、权益乘数、流动资产/总资产、流动负债/负债合计

表 3 不同财务预警模型的预测结果

公司编号	模型	SVM 模型		Logistic 模型		信息融合模型	
	标签	概率	标签	概率	标签	概率	
公司 1	1	0.8915	1	0.9556	1	0.9944	
公司 2	0	0.9667	0	1.0000	0	1.0000	
公司 3	0	0.8463	0	0.9988	0	0.9998	
公司 4	1	0.8947	1	1.0000	1	1.0000	
公司 5	1	0.9273	1	0.8968	1	0.9911	
公司 6	0	0.5977	0	0.9893	0	0.9928	
公司 7	0	0.5353	0	0.9878	0	0.9894	
公司 8	1	0.5120	1	0.8216	1	0.8286	
公司 9	0	0.9432	0	0.9999	0	1.0000	
公司 10	0	0.5819	0	0.9999	0	0.9999	
公司 11	0	0.8482	0	1.0000	0	1.0000	
公司 12	1	0.6643	1	0.9419	1	0.9698	
公司 13	0	0.9434	0	0.9979	0	0.9999	
公司 14	0	0.7068	1	0.5261	0	0.6847	
公司 15	0	0.7946	0	0.9700	0	0.9921	
公司 16	0	0.8289	0	0.9998	0	1.0000	
公司 17	1	0.5421	0	0.8598	0	0.8382	
公司 18	0	0.6190	0	0.9310	0	0.9564	
公司 19	0	0.7435	0	0.9857	0	0.9950	
公司 20	1	0.6265	1	0.5948	1	0.7112	
公司 21	0	0.9324	0	0.9999	0	1.0000	

表 4 不同财务预警模型的预测准确率

结果	模型	SVM 模型	Logistic 模型	信息融合模型
准确率		90.48%	90.48%	95.24%
错误率		9.52%	9.52%	4.76%

从表 4.3 和 4.4 中可以看出,SVM 和 Logistic 模型在分类问题中具有很好的效果,都能达到较高的分类准确率。在预测的结果中,可以清楚的观察到,当 SVM 和 Logistic 模型的预测结果相同时,信息融合算法给出的预测结果也相同,而当二者模型的预测结果不同,信息融合技术可以通过基于概率的证据理论推算方法计算出最终的预测结果,并且将预测准确率从 90.48% 提高到 95.24%。信息融合的优势就是可以将不同信息源传递来的数据以一种可信用度逻辑推理方法计算出最终的结果。基于信息融合的数据挖掘模型比单一的数据挖掘方法得到结果的准确率要高,而且该模型还保持了不同数据挖掘模型各自的优势。该实证分析同时证明了信息融合技术在数据挖掘结果的“二次”处理上具有一定的有效性和可行性。

4 结语

数据挖掘模型为企业财务预警提供一种较新的有效方法,并且已经被广泛的运用与财务预警的研究中。SVM 和 Logistic 模型也是较好解决分类问

题的主流数据挖掘方法,但是不同的数据挖掘方法有时会产生不同的结果,“一次”挖掘得到的结果有时不能很好的提供决策支持。信息融合技术通过基于可信用度逻辑推理的算法将不同数据挖掘得到的结果进行“二次”计算的过程,并且得到了更加可靠的结果用于决策支持,同时也提高了决策的安全性。

文中分别利用 SVM 和 Logistic 回归建立了财务预警模型,预测准确率均达到 90.48%,再利用信息融合技术后将预测准确率提高到 95.24%,证明了信息融合技术能够有效提升数据挖掘结果,并且解决了两种数据挖掘结果不一致问题。本文利用我国制造业上市公司的财务数据建立基于信息融合的数据挖掘模型,提高了预测准确率,证明该方法的有效性和可行性,也为公司财务预警提供了较为可靠的方法。

参考文献:

[1] Altman E I, Hanldeman R G, Narayanan P. ZETA analysis: A new model to identify bankruptcy risk of corporations[J]. Journal of Banking and Finance, 1977, 1 (1):29-54.

[2] Ohlson J A. Financial ratios and the probabilistic prediction of bankruptcy[J]. Journal of Accounting Research, 1980,18(19):109-131.

[3] Odom M, Sharda R. A neural network model for bankruptcy prediction[J]. Neural Networks in Finance and Investing,1993,2:177-186.

[4] 杨淑娥,王乐平. 基于 BP 神经网络和面板数据的上市公司财务危机预警[J]. 系统工程理论与实践,2007,2 (6):61-67.

[5] 姚靠华,蒋艳辉. 基于决策树的财务预警[J]. 系统工程,2005,23(7):102-106.

[6] Gestel T, Baesens B, Suykens J, et al. Bayesian kernel based classification for financial distress detection[J]. European Journal of Operational Research, 2006, 172 (3):979-1003.

[7] Shin K, Lee T, Kim H. An application of support vector machine in bankruptcy prediction model [J]. Expert Systems with Applications,2005,28(1):127-135.

[8] 杨毓,蒙肖莲. 用支持向量机(SVM)构建企业破产预测模型[J]. 金融研究,2006,(10):65-75.

[9] Koyuncugil A S, Ozgulba N, Financial early warning system model and data mining application for risk detection[J]. Expert Systems with Applications, 2012, 39 (6):6238-6253.

[10] Mina J H, Lee Y C. Bankruptcy prediction using support vector machine with optimal choice of kernel func-

- tion parameters[J]. Expert Systems with Applications, 2005, 28(4): 603-614.
- [11] 杨淑娥, 黄礼. 基于 BP 神经网络的上市公司财务预警模型[J]. 系统工程理论与实践, 2005, 25(1): 12-18.
- [12] 康晓玲, 张懿. 企业财务预警模型研究—基于中报数据与可持续增长模型的重构[J]. 科研管理, 2009, 30(1): 45-64.
- [13] 谭久均. 财务指标与违约距离相融合的上市公司财务预警模型[J]. 系统工程, 2005, 23(9): 111-117.
- [14] 孙洁, 李辉, 韩建光. 基于滚动时间窗口支持向量机的财务困境预测动态建模[J]. 管理工程学报, 2010, 24(4): 174-180.
- [15] 韩建光, 惠晓峰, 孙洁. 基于多特征子集组合分类器的企业财务困境预测[J]. 系统管理学报, 2010, 19(4): 420-427.
- [16] 刘遵雄, 郑淑娟, 秦宾, 等. L1 正则化 Logistic 回归在财务预警中的应用[J]. 经济数学, 2012(2): 106-110.
- [17] 周宗放, 肖珉. 我国企业集团上市公司财务预警与信用风险评估研究[D]. 四川: 电子科技大学, 2012.
- [18] 孟杰. 随机森林模型在财务失败预警中的应用[J]. 统计与决策, 2014, (4): 179-181.
- [19] 辛金国, 关建清. 基于数据挖掘民营上市公司绩效评价研究探索[J]. 中国管理科学, 2012, 20(57): 114-119.
- [20] 杨海珍, 荆中博, 魏先华, 等. 银行破产的财务因素分析: 金融危机冲击下美国银行业的实证[J]. 中国管理科学, 2012, 20(1): 71-77.
- [21] 文柯. 基于 Logistic 的上市公司产融结合风险预警模型研究[J]. 中国管理科学, 2012, 20(S2): 346-350.

Based on Information Fusion Technique with Data Mining in the Application of Finance Early-Warning

ZHANG Liang^{1,2}, ZHANG Ling-ling^{1,2,3}, CHEN Yi-bing^{2,4}, TENG Wei-li⁵

(1. School of Management, University of Chinese Academy of Sciences, Beijing 100190, China;

2. Research Center on Fictitious Economy and Data Science, Chinese Academy of Science, Beijing 100190, China;

3. Key Laboratory of Big Data Mining and Knowledge Management, CAS, Beijing 100190, China;

4. Institute of Policy and Management, Chinese Academy of Science, Beijing 100190, China;

5. Nottingham Business School Doctorate of Business Administration, Nottingham Burton Street, NG1 4BU)

Abstract: Different data mining methods for classification can produce different results. However, “one time” data mining process cannot often obtain a well support decision, so we introduce information fusion technique to fuse the different results to gain an optimal solution. In this paper, information fusion technique is used to build a finance early-warning model based on data mining methods such as SVM and Logistic model, which can integrate the respective strengths from different data mining methods to improve the prediction accuracy rate, it fuses the different data mining results to gain the prediction results for reliable decision. The real dataset of Chinese listed manufacturing companies is selected to predict the finance risk with information fusion technique based on SVM and Logistic model, and a higher prediction accuracy than those of the two methods respectively is obtained.

Key words: finance early-warning; SVM model; logistic model; information fusion