

上市公司财务困境预测

——基于ANFIS和马氏距离判别法的组合模型

李力(博士生导师),冯涛

(哈尔滨工业大学深圳研究生院,深圳 518055)

【摘要】财务困境预测是现代金融领域的一个重要研究方向,人工智能技术的发展为该研究提供了更多方法。鉴于单一预测模型的缺陷与低准确率,本文首先运用基于分类的神经网络模糊推理系统、马氏距离判别法和贝叶斯网络构建组合预测模型。然后,以我国制造业类上市公司为研究对象进行财务困境预测实证研究。研究结果表明,该组合预测模型克服了单一预测模型的一些缺陷,对ST公司的预测准确率明显高于其他几种模型,具有较好的预测性能。

【关键词】财务困境预测;自适应神经网络模糊推理系统;马氏距离判别法;贝叶斯分类器

一、引言

企业财务困境预测是一项基于企业已有的财务或非财务方面的信息,运用数理统计和人工智能等方法,预测企业在未来是否陷入财务困境的研究。企业财务困境也称为企业破产、财务危机、财务失败等。企业财务困境预测研究对于企业经营者、投资者和政府监管部门等具有重要的现实意义。

国外财务困境预测研究最早兴起于20世纪60年代,当时的学者使用的都是一些比较简单的数理统计方法。Beaver(1966)最初使用的是一元判别分析,而Altman(1968)使用的是多元判别分析。随着计算机技术的迅速发展,在21世纪初的时候,人工智能方法便开始被应用到企业财务困境预测领域。

吴德胜等(2004)比较了神经网络模型和多元线性判别的财务困境预测能力,发现多元线性判别在二分类问题上具有较大优越性,而神经网络在多分类问题上具有较高的准确率。杨淑娥等(2005)使用BP神经网络,发现BP神经网络的预测准确率相比用主成分分析构建的模型有很大提高。

学者Shin(2005)和Min等(2005)使用支持向量机预测企业的财务困境。Setiono(2011)等在预测企业财务困境时提出用一种新方法训练最小神经网络,它可以产生简单易理解的分​​类规则。

最近几年来,有些学者又不满足于单一的预测模型,他们开始使用组合预测模型,通过对不同预测模型的取长补短,使得组合预测模型的预测性能得到进一步提升。时建中等(2013)用分类器集成技术,以BP神经网络为分类学习算法,建立了基于RS-Bag算法的神经网络分类器

模型。他们的研究表明,其预测精度和泛化性能由于一般神经网络分类器,也优于Bagging分类器集成和RS分类器集成。

为了克服神经网络黑箱的缺点,Chen(2013)采用粒子群优化算法(PSO)为减法聚类(SC)获得适合的参数,再整合自适应神经网络模糊推理系统建立预测企业财务困境的模型。实证结果表明,与普通的ANFIS和PSO-ANFIS相比,他的模型PS-ANFIS结果产生的平均绝对百分比误差和均方根误差要更小。

二、研究模型理论

1. 基于分类的ANFIS模型。由于传统的时间序列模型只能适用于单一变量,而传统的人工神经网络是一个黑箱,不能表达模糊语言,不能很好地表达人脑的推理功能,因此模糊推理本身不具备学习功能。为了克服以上这些问题,本文选用自适应神经网络模糊推理系统,它将人工神经网络和模糊推理相结合,既发挥了二者的优点,又弥补了各自的不足。此外,本文还在传统的ANFIS基础上,增加了可对其结果进行分类的功能,模型可根据结果直接判别企业的财务状况类别。

基于分类的ANFIS模型的运行原理是先用训练样本数据训练模型参数,即对于一个给定的输入输出数据集,通过减法聚类算法创建一个模糊推理系统,供ANFIS进行迭代训练。它的隶属度函数通过反向传播算法和最小二乘算法的组合进行调整,这种调整可以允许模糊系统从建模的数据中学习。由于本文研究的是一个二分类问题,即判断企业是财务困境企业还是财务健康企业,因此基于分类的ANFIS模型可以根据其输出结果直接进行判别分类。

ANFIS模型是由Jang提出的一种多层次且结合模糊推理和神经网络的产物。它首先通过初始模糊推理系统获得隶属度函数个数和规则个数,然后采用最小方差估计得到推理规则结论部分的参数,最后经过多次迭代训练之后获得判别模型。

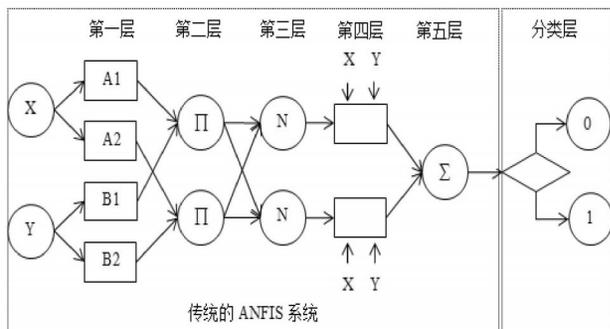


图1 基于分类的ANFIS模型结构

图1中,左边虚线框中的流程图是传统的ANFIS系统。图中第k层的第i个位置的节点用 $O_{k,i}$ 来表示;同一层的节点函数属于同一种函数类型。

第一层是模糊化层,该层的每个节点i是以节点函数表示的方形节点,见式(1)。 $O_{1,i}$ 是 A_i 和 B_i 的隶属度函数, A_i 和 B_i 是与该节点函数值相关的语言变量。

$$O_{1,i} = \mu A_i(x_i), i=1, 2; O_{1,i} = \mu B_j(y_j), j=1, 2 \quad (1)$$

第二层是模糊推理层,该层的节点在图1中用 Π 表示,将输入信号相乘,并将其乘积结果输出为:

$$O_{2,i} = w_i = \mu A_i(x) \times \mu B_j(y), i=1, 2 \quad (2)$$

该层的每个节点代表一条模糊规则,它是用于匹配模糊规则的前提,首先计算出每个规则的适用度 ω_i ,然后通过模糊集合的运算,将其中最大的作为权值。

第三层是归一化层,将所有规则的可信度归一化。在图1中用N表示,第i个节点计算第i条规则的 ω_i 与全部规则 ω 值之和的比值如下:

$$O_{3,i} = \bar{\omega}_i = \frac{\omega_i}{\omega_1 + \omega_2} \quad i=1, 2 \quad (3)$$

第四层是规则生成层,表示其中某条规则对应的结果部分。该层每个节点i为自适应节点,其输出为:

$$O_{4,i} = \bar{\omega}_i f_i = \bar{\omega}_i (p_1 x_1 + q_1 x_2 + r_1) \quad (4)$$

其中, p_1, q_1, r_1 是参数。

第五层是解模糊层,它的作用是计算模糊系统的输出值,即将每条规则的输出进行结果加权平均。该节点用 Σ 表示,其总输出为:

$$O_{5,i} = \sum_j \bar{\omega}_j f_j = \frac{\sum_{i=1}^n \omega_i f_i}{\sum_{i=1}^n \omega_i} \quad (5)$$

本文将上市公司的财务状况分为两个类别,并用数值来标记企业财务状况,财务困境公司标为0,财务健康

公司标为1。由于ANFIS的输出是拟合后的结果,因此它的输出值会接近实际值,分布在区间 $[0 \sim 1]$ 的附近。基于分类的ANFIS系统会在传统的ANFIS系统后面加入一个分类层,该分类层可以根据事先设定的分割值来判断企业的财务状况。图1中右边虚线框中的流程图是本文新设计的对传统ANFI模型结果的分层。

2. 马氏距离判别法。马氏距离是由印度统计学家Mahalanobis提出的一种计算广义距离的方法。

设 Σ 为指标的协方差矩阵,即:

$$\Sigma = (\sigma_{ij})_{p \times p} \quad (6)$$

$$\text{其中, } \sigma_{ij} = \frac{1}{n-1} \sum_{a=1}^n (x_{ai} - \bar{x}_i)(x_{aj} - \bar{x}_j), i, j=1, \dots, p$$

$$\bar{x}_i = \frac{1}{n-1} \sum_{a=1}^n x_{ai} \quad \bar{x}_j = \frac{1}{n-1} \sum_{a=1}^n x_{aj}$$

假设 x 是来自协方差矩阵为 Σ 、均值向量为 μ 的总体G内的一个行向量,那么:

$$d(x, G) = \sqrt{(x - \mu) \Sigma^{-1} (x - \mu)^T} \quad (7)$$

d 就是n维向量 x 与总体G的马氏距离。

设有两个总体 G_1 和 G_2 ,对于某个样本 X ,要判断它属于哪一个总体,通常的做法是计算样本 X 分别到 G_1 和 G_2 的马氏距离 $D^2(X, G_1)$ 和 $D^2(X, G_2)$,并按照如下的判别规则来进行判断:

$$\begin{cases} X \in G_1, \text{ 如果 } D^2(X, G_1) \leq D^2(X, G_2) \\ X \in G_2, \text{ 如果 } D^2(X, G_1) > D^2(X, G_2) \end{cases}$$

与欧氏距离相比,马氏距离主要有以下几个优点:第一,量纲对它没有影响,任意两点间的马氏距离与原来数据的单位是没有关系的;第二,无论是使用中心化数据,还是使用标准化数据(即原始数据与均值之差),两点间的马氏距离计算结果是相同的;第三,马氏距离还能排除各变量之间的相关性干扰,这点可以从其公式中的协方差矩阵看出来。

3. 朴素贝叶斯分类器。贝叶斯网络是一类图形化网络,它主要是基于概率推理展开的,而贝叶斯公式就是整个网络的基础。所谓概率推理,就是通过某些变量的信息来获得其他概率信息的过程。贝叶斯网络是由学者Pearl在1988年提出的,它现在主要运用于推理和不确定知识表达领域。

设 (B_1, B_2, \dots, B_n) 是E的一个互不相容的完备事件组,且 $P(B_i) > 0$,A是E的任意事件,则有:

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^n P(B_j)P(A|B_j)} \quad (8)$$

式(8)即为贝叶斯公式。

当给定类变量时,属性变量之间条件独立,即满足式(9)时,该条件称为朴素贝叶斯假设。根据朴素贝叶斯假设,式(10)即可成立,其中 α 是与C无关的规范化常数。

$$P(x_1, \dots, x_n | c) = \prod_{i=1}^n P(x_i | c) \quad (9)$$

$$P(c | x_1, \dots, x_n) = \frac{P(c, x_1, \dots, x_n)}{P(x_1, \dots, x_n)} = \frac{P(c)P(x_1, \dots, x_n)}{P(x_1, \dots, x_n)} = \frac{P(c) \prod_{i=1}^n P(x_i | c)}{P(x_1, \dots, x_n)} = \alpha P(c) \prod_{i=1}^n P(x_i | c) \quad (10)$$

对概率分布 $P(X_1, \dots, X_n, C)$,使用式(11)进行预测:

$$\arg \max_{(c, x_1, \dots, x_n)} \left\{ P(c) \prod_{i=1}^n P(x_i | c) \right\} \quad (11)$$

对变量C进行预测的分类器称之为朴素贝叶斯分类器。

设事件C为上市公司会陷入财务困境,事件 X_1 为基于分类的ANFIS模型对事件C中个体的预测结果,事件 X_2 为马氏距离判别法对事件C中个体的预测结果。由于只涉及基于分类的ANFIS和马氏距离判别法的预测结果,所以式(10)只需要考虑当 $n=2$ 时的情形。 $P(X_1|C)$ 和 $P(X_2|C)$ 分别是基于分类的ANFIS模型和马氏距离判别法对ST公司和非ST公司的判别准确率。 $P(X_1, X_2)$ 可以视为一个与事件C无关的规范化常数 α ,然后就可以使用公式(10)计算个体分别是ST公司和非ST公司的概率,最后比较两个概率值的大小,哪个概率值大就认为个体属于该种类型的公司。本文所有模型实验都在MATLAB R2011b上编程实现。

三、实证研究

1. 数据搜集与预处理。本文根据国内学者的普遍做法,将上市公司因财务状况异常而被ST作为企业陷入财务困境的标志,把ST公司作为财务困境公司,把非ST公司作为财务健康公司。

本文选择的ST公司是在2008~2013年间沪深交易所中因财务状况异常而被特别处理的A股制造业中的95家上市公司。之所以选择制造业的上市公司,是因为制造业上市公司数量相对比较多,其发展也比较稳定。同时,按照ST公司与非ST公司1:2的比例随机选择同行业的190家财务健康公司作为配对样本,配对样本要求对应的ST公司在被特别处理的当年是财务健康公司。同时,将搜集的上市公司样本以7:3的比例分为两个子样本,分别为训练样本和检验样本。

在分配训练样本与检验样本时,还考虑了不同会计

年度对上市公司财务状况的影响,所以两个子样本中同一会计年度的样本数量也按照7:3的比例分类。训练样本中的数据是用来训练模型的参数,而检验样本中的数据则只用来检验模型的预测准确率,检验样本不在模型训练的过程中使用。

假定上市公司是在T年被特别处理,本文使用(T-2)会计年度的财务数据来研究企业财务困境预测,即提前两年预测企业是否会陷入财务困境。刘彦文(2009)统计了2002~2008年国内外发表的118篇财务困境预测研究相关文献,列出了在国内外相关研究中均对最终预测模型产生显著贡献的预测变量及其出现频次。本文根据刘彦文(2009)所统计的预测指标频次统计表,把其中频次超过50次的财务指标作为本文的模型初始自变量,主要有22个财务指标,具体见表1。

表 1 初始自变量

种类	财务指标
盈利能力	总资产净利率*、销售净利率、净资产收益率*
偿债能力	流动比率、速动比率、资产负债率*、经营性现金流量流动负债比、经营性现金流量债务总额比*
经营能力	营运资本与总资产比率、应收账款周转率、总资产周转率、流动资产周转率、应收账款与流动资产比率、其他应收账款与流动资产比率、营业利润比营业总收入、应收账款比销售收入*
成长能力	营业收入同比增长率、净资产同比增长率
其他	留存收益比总资产、销售现金比率、资产现金回收率、非经常性损益与利润总额比率

一般认为,模型中加入的变量越多,其预测准确率会越高。但是太多的变量会增加模型的复杂度,而且会限制模型的适用性。同时,模型使用太多的变量可能会出现过度拟合的情况,对于训练样本分类会比较成功,但是未必会对训练样本外的其他分类有效。另外,考虑到ANFIS模型对于每个输入变量都有模糊推理过程,模型结构会因变量的增多而更加复杂,机器运算速度也会相应减慢,因此,为了降低模型的输入变量维度,同时为了提高预测模型的准确性,本文将对22个初始自变量进行预处理,剔除一些不太重要的财务指标。

本文采用逐步判别法筛选最终自变量。本文使用统计软件SPSS20.0对最初获得的22个初始财务自变量进行了逐步判别分析。软件的统计结果如表2所示。

表 2 逐步判别分析处理结果

判别函数	特征值	方差(%)	累计方差(%)	正则相关性
1	1.201 ^a	100.0	100.0	0.739

通过这22个初始财务指标最终提取到一个判别函数,该判别函数可以解释100%的方差,说明判别函数最终提取的5个财务变量与上市公司是否会被ST具有较强的相关性。

同时,这5个财务变量都是在显著性水平5%下取得的,所以可以通过这5个财务变量来预测上市公司未来的财务状况。5个财务变量见表1中带“*”的指标。

2. 基于分类的ANFIS预测。首先,把训练样本组中公司的财务数据以及财务状况类别(0和1)分别作为输入和输出数据,建立初始模糊推理系统来训练模型参数。通过试错法来确定ANFIS模型中函数genfis2中的聚类半径。当聚类半径确定为0.45时,模型根据输入数据获得的拟合结果与实际结果最为接近。

然后,使用自适应模糊神经网络的优化能力以提高预测系统的性能。本文选择模型的训练迭代次数为1000次,使用之前用减法聚类算法初始化后的模糊推理系统和训练样本中的财务数据来训练ANFIS系统。

图2显示的是经过训练样本数据进行参数训练之后的ANFIS模型对检验样本数据的模拟输出结果,横轴表示样本数量,纵轴表示上市公司的财务状况(0表示公司为ST,1表示公司为非ST),加号表示样本的实际结果,圆圈表示模型以检验样本数据为输入数据的模拟结果。从图2中可以看到,虽然模型的模拟结果并没有与实际值重叠在一起,但是大部分的模拟结果都接近其实际值,只有个别的模拟结果偏离其实际值较远,说明训练之后的该模型可以模拟大部分的样本。

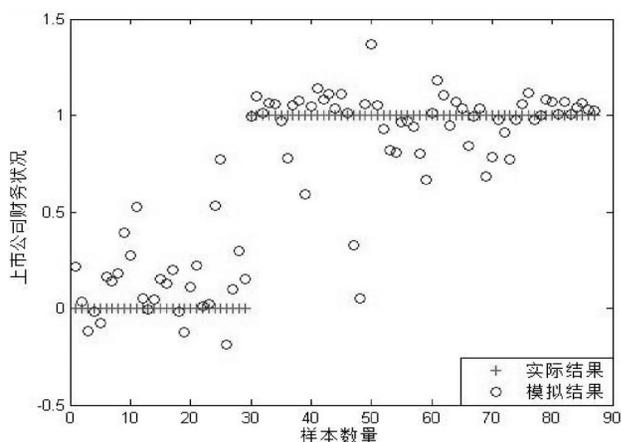


图2 ANFIS对检验样本数据的模拟结果

最后,基于分类的ANFIS模型的分类层将对上一步骤中模型的模拟结果进行分类。在保证判别分类误差率最小的前提下,本文设定分割值为0.5,即输出值小于0.5的判别为ST公司,而大于0.5的判别为非ST公司。基于分类的ANFIS模型最终输出结果如图3所示,从图中可以非常直观地看到,有5个检验样本被模型误判,而其余的检验样本都判断正确。

检验样本的综合预测准确率达到94.25%,说明该模型的预测效果较好。同时,该模型对于ST公司和非ST公司两种财务状况类别的预测判别准确率分别是89.66%和96.55%,说明模型对财务健康企业的预测能力更强。

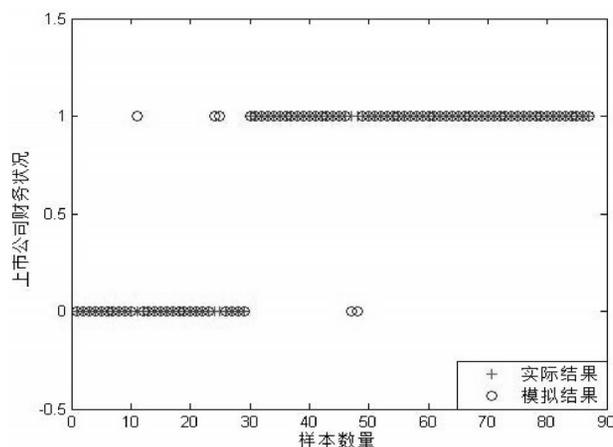


图3 基于分类的ANFIS对检验样本数据的模拟结果

3. 马氏距离判别法预测。在运用马氏距离判别法预测企业财务困境时,将训练样本中的ST公司和非ST公司分为两个子样本,用这两个子样本来表示ST公司和非ST公司在空间中分布的两个类别,记为 M_{ST} 和 M_{NST} 。记检验样本中公司个体为 C ,然后计算每一个个体的 $D^2(C, M_{ST})$ 和 $D^2(C, M_{NST})$ 。若 $D^2(C, M_{ST})$ 小于 $D^2(C, M_{NST})$,则 C 公司判别为ST公司;反之, C 公司为非ST公司。为方便起见,当同一个体的两个马氏距离相同时,判别为ST公司。马氏距离判别法关于ST公司与非ST公司的预测准确率分别为96.55%和89.66%,综合预测准确率为91.95%。可见,马氏距离判别法对ST和非ST公司的判别精确率都较高,所以用该方法来判别上市公司的财务状况是非常有效的,从而也直接证明以本文得到的五个财务数据作为空间向量的上市公司确实存在聚类现象。同时,与基于分类的ANFIS模型相反,马氏距离判别法对ST公司识别更加准确。

4. 组合模型预测。在用朴素贝叶斯分类器对结果进行组合预测时,首先需要知道事件 C 的先验概率 $P(C)$ 。由于因其他情况(非财务状况异常)被ST的上市公司数量较少,本文采用2010~2013年因财务状况异常而被ST的A股上市公司占总的A股数量的平均比例作为事件 C 的先验概率 $P(C)$ 。

本文统计了2010~2013年四年间A股中因财务状况异常而被ST的上市公司数量及当年A股中所有上市公司的数量,然后计算ST公司占A股上市公司的平均百分比为1.07%,即 $P(C)$ 为0.0107。 $P(X_1|C)$ 和 $P(X_2|C)$ 的值就是前文基于分类的ANFIS模型和马氏距离判别法对ST公司和非ST公司的预测准确率,这样就可使用组合模型来预测上市公司未来的财务情况。

表3是组合模型分别关于训练样本和检验样本的预测准确率。模型对训练样本的预测准确率是用训练样本数据对经过训练之后的模型进行回代的结果。虽然组合模型对训练样本和检验样本都具有较高的预测准确率,

但是训练样本数据会高估模型的预测准确率,所以只需考虑组合模型对检验样本的预测准确率。组合模型对ST公司的准确率接近90%,而对非ST公司的预测准确率高达96.55%,综合预测准确率也接近95%,说明组合模型的预测性能非常好。

表 3 模型关于样本的预测正确率

组 别	ST 预测 准确率(%)	非ST 预测 准确率(%)	综合预测 准确率(%)
训练样本	92.42	89.39	90.40
检验样本	89.66	96.55	94.25

5. 模型稳健性检验。为了增强实证研究分析的稳健性,本文采用了交叉验证的方法,分五次分别从ST公司和NST公司样本中各抽取29家和58家上市公司组成检验样本,并将剩下的ST公司和NST公司共198家上市公司作为训练样本。在用训练样本的数据建立模型之后,再用检验样本检验组合模型预测上市公司财务困境的正确率。表4中第二行是已经在前文中详细说明的例子,其余四项省略了计算过程。

表 4 组合预测模型稳健性检验结果

模 型	ST 预测 准确率(%)	NST 预测 准确率(%)	综合 准确率(%)
组合模型(1)	89.66	96.55	94.25
组合模型(2)	89.66	96.55	94.25
组合模型(3)	89.66	98.28	95.40
组合模型(4)	86.21	98.28	94.25
组合模型(5)	82.76	96.55	91.95
组合模型(平均)	87.59	97.24	94.02
神经网络	63.64	95.12	-
支持向量机	77.27	95.12	-
Logit 模型	54.55	92.68	-

如表4所示,无论是对于ST公司还是非ST公司,其预测准确率都比较相近,不存在偶然性的问题。组合模型对ST公司的平均预测准确率为87.59%,对非ST公司的平均预测准确率为97.24%,其综合预测准确率为94.02%。由于模型对ST公司和NST公司的预测准确率不相等,因此该综合准确率会根据ST公司与NST公司的比例变化而变化。使用企业财务困境预测准确率和财务健康预测准确率更加符合实际,也易于比较各个模型之间的预测性能。因此,在比较本文的组合模型与神经网络、支持向量机和Logit模型时只考虑分别预测ST与非ST公司的准确率,不考虑综合准确率。从表4中可以看出,虽然其他三个模型对非ST公司的预测准确率没有本文组合模型高,但也非常接近;其他三个模型对ST公司的预测准确率则远远不如本文的组合模型。

四、结论

本文构建了一个基于三种算法的企业财务困境预测模型,首先分别使用基于分类的ANFIS系统和马氏距离判别法,根据企业财务数据对企业未来的财务状况进行判别预测,再将两者的结果作为贝叶斯判别法的输入数据,最后得出综合的预测结果。

实证结果表明,ANFIS用了模糊的if-then规则构建了人类知识的定性方面,具有较强的预测性能,而最后的分类层可以直接根据数据对样本进行分类。马氏距离判别法的应用证明:如果把某些财务数据作为ST公司和非ST公司在空间中的向量,表明ST公司和非ST公司在空间中会存在聚类现象,能更加直观地来判别ST与非ST公司。应用基于分类的ANFIS模型和马氏距离判别法两种不同运行原理的判别法,再使用朴素贝叶斯分类法可以提高整个模型的预测能力,其对ST公司的预测准确率明显高于其他几个预测模型。

主要参考文献

Beaver W. H.. Financial Ratios as Predictors of Failure [J]. Journal of Accounting Research, 1966(7).

Frydman H., Altman E.I, Kao D.. Introducing Recursive Partitioning for Financial Classification: the Case of Financial Distress [J]. Journal of Finance, 1985(1).

吴德胜等.不同模型在财务预警实证中的比较研究[J].管理工程学报,2004(2).

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(1).

Shin K. S., Lee T. S., Kim H. J.. An Application of Support Vector Machines in Bankruptcy Prediction Model [J]. Expert Systems with Applications, 2005(28).

Min J. H., Lee Y. C.. Bankruptcy Prediction Using Support Vector Machine with Optimal Choice of Kernel Function Parameters [J]. Expert Systems with Applications, 2005(28).

时建中,程龙生,牛俊磊.基于RS-Bag分类器集成技术的上市公司财务危机预测[J].数理统计与管理,2013(5).

肖泰琨,高嵩.贝叶斯网络在智能信息处理中的应用[M].北京:国防工业出版社,2012.

刘彦文.上市公司财务危机预警模型研究[D].大连:大连理工大学,2009.

【基金项目】教育部人文社会科学研究规划基金项目“碳信息披露对企业价值创造的影响研究”(项目编号:13YJAZH044);国家自然科学基金项目“粒计算拓展模型及多知识库集成信息融合研究”(项目编号:61173052)