

基于违约区分程度最大权重的信用评级模型^①

周 颖

(大连理工大学经济管理学院, 大连 116024)

摘要: 信用评级方程是信用评级模型的核心, 权重是影响信用评级方程评价准确度的重要因素, 合理的权重对于信用评级至关重要。通过以违约与非违约客户信用得分单位离散度的平均距离最大为目标函数, 构建非线性规划模型的思路, 反推各指标的最优权重组合。通过 PR 曲线方法对信用评级模型的合理性进行检验, 并对 1 231 个小企业贷款进行了信用评级。通过构建信用评级方程 $S=S(W)$ 的权重向量 W 与违约与否的两类客户平均信用得分的距离 D 之间的函数关系, 以违约与否的两类客户平均信用得分距离 D 最大为目标建立非线性规划, 反推一组最优权重 W 。保证了信用评级方程的评级结果, 能够最大限度地区分违约与非违约客户。改变了现有研究的权重向量与整体违约鉴别能力无关的弊端。非财务指标相对于财务指标在小企业评级中更重要, 权重高达 58%。非财务指标中“城市居民人均可支配收入”指标的权重最大, 为 14.7%。通过 PR 曲线方法验证了其合理性。研究表明, “基于违约状态区分程度最大权重的小企业信用评级模型”, 对客户违约与否的区分精度, 优于现有研究的同类经典模型。

关键词: 信用评级; 权重向量; 最优权重; 最优赋权; 小企业评级

中图分类号: F830.56; O221.2 文献标识码: A 文章编号: 1007-9807(2019)09-0052-14

0 引 言

信用评级是对一个企业或个人的违约可能性进行评估。信用评级的结果直接影响金融机构进行贷款决策、贷款定价、债券定价。信用评级不准会导致严重的后果, 小到因为贷款决策和贷款定价错误给银行造成损失, 大到引发全球性的金融危机。例如, 穆迪和标准普尔给次级抵押贷款评以高信用等级, 最终导致市场崩溃, 引起了 2008 年美国金融危机^[1]。

信用评级方程中信用评价指标的权重变化直接影响信用评级模型的评级结果, 合理的权重才能保证信用评级模型的评级准确性。因此研究如何为信用评价指标赋予合理的权重, 保证信用评级模型能够准确区分违约与非违约客户具有重大

意义。

应该指出, 信用评价指标选择和信用评级方程中的赋权是相辅相成的。指标不同, 评级结果会截然不同。对相同的指标赋予不同的权重, 结果也会大相径庭。因此指标的遴选和权重确定都是信用评级研究中重要问题。本文研究信用评级方程中最优权重的确定。

现有的信用指标赋权方法主要分 3 大类: 基于主客观结合的组合赋权法、基于回归分析的赋权法和基于人工智能的赋权法。

1) 基于主客观结合的组合赋权

主客观组合赋权是将主观赋权与客观赋权结合的赋权方法, 它结合了两种赋权方法的优点, 兼顾了赋权的主观偏好和客观数据。Che 等^[2]通过模糊层次分析法(FAHP)和数据包络法(DEA)结

① 收稿日期: 2017-11-03; 修订日期: 2018-11-14。

基金项目: 国家自然科学基金重点项目(71731003; 71431002); 爱德力智能科技(厦门)有限公司智能风险管控模型与算法项目(2019-01)。

作者简介: 周 颖(1966—), 女, 吉林长春人, 博士, 副教授, 硕士生导师。zhouying@dlut.edu.cn

合的方法进行组合赋权,建立了台湾中小企业的信用评级模型。张目等^[3]应用相对熵和可变模糊集理论进行组合赋权,构建了针对战略性新兴产业企业信用评价模型。陈晓红和杨志慧^[4]将基于群决策的AHP主观赋权方法与基于因子分析(FA)的客观赋权方法相结合,提出了改进模糊综合评价模型(AFF),对我国中小企业进行评级。李步军等^[5]结合主客观赋权方法的优点,采用灰区间关联分析对信用评价指标进行赋权,建立了个人信用优化评估模型。

组合赋权法虽然兼顾了主客观因素进行赋权,但只是根据单个指标的重要程度进行赋权,并未考虑信用评价结果对客户违约状态的鉴别能力,仍存在局限性。

2) 基于回归分析的赋权

基于回归分析对信用指标赋权,典型的方法有3种。第一种是判别分析法(discriminant analysis),最早基于判别分析法赋权的信用评级模型是由Altman^[6]提出的著名的Z-score模型,该模型是5个财务指标的线性加权,之后Desai等^[7]、王春峰和李文华^[8]也通过判别分析法赋权建立信用评级模型。第二种是Probit回归法(probit regression),Grabrowsky和Talley^[9]最先将Probit回归法用于信用评级模型的赋权。Duca和Whitesell^[10]通过Probit回归赋权建立了信用评级方程,对个人信用卡业务进行评级。第三种是逻辑回归法(logistic regression),Hand和Henley^[11]通过了逻辑回归法对信用指标进行赋权,构建了消费贷款的信用评级模型。李战江^[12]将逻辑回归法与统计抽样理论中的分层思想相结合对指标进行赋权,构建了小企业信用评级模型。

基于回归分析的赋权法,虽然在区分违约与非违约的精确度上优于组合赋权法,但回归法只是对指标数据和违约状态的拟合,仍然无法避免赋权后计算所得的信用得分存在大量交叠的现象。

3) 基于人工智能的赋权

第一类基于人工智能的赋权法是支持向量机法。程砚秋^[13]采用不均支持向量机法,通过对违约状态影响越大、评价指标权重越大的赋权思路,对信用评价指标进行赋权,构建了小企业信用风险评价模型。

第二类基于人工智能的赋权法是神经网络法

(neural network)。West^[14]对比了5种神经网络法的赋权效果,进而构建信用评级模型,并将神经网络法与判别分析法、逻辑回归法、K-近邻法、核密度估计法、决策树法的实证结果进行对比,发现径向基函数神经网络法更适用于实践中构建信用评级模型。庞素琳^[15]运用概率神经网络(PNN)法对中国106家上市公司进行了信用评价和预警分析。Tsai和Wu^[16]采用了多层感知器神经网络法赋权,构建信用评级模型。

第三类基于人工智能的是遗传规划算法(genetic programming)。王春峰和康莉^[17]用遗传规划算法对130家中国企业进行信用评价,发现用遗传规划算法构建的模型在预测精度、实用价值以及鲁棒性方面都优于传统的统计模型、神经网络模型以及决策树模型。Ong等^[18]用遗传规划算法构建了信用评分模型,研究表明遗传规划算法构建的模型在预测方面优于人工神经网络法、粗糙集法、决策树法和逻辑回归法。Abdou^[19]基于埃及公共部门银行的数据,将遗传规划算法建立的信用评分模型同Probit分析、证据权重法建立的模型进行对比,发现遗传规划算法相对于其他方法具有更高的平均正确分类率(ACC)以及更低的错判成本(EMC)。

现有研究^[2-19]在对信用评价指标赋权,构建信用评级方程时,大多未能考虑信用得分对违约与非违约客户的区分,且不能保证信用得分能够有效区分违约和非违约客户,导致大量非违约与违约客户的信用得分发生交叠,非违约客户的信用得分不能显著高于、甚至低于违约客户的信用得分。

针对上述问题,本研究以违约与非违约客户区分程度最大为目标函数,构建非线性规划模型,反推一组最优的指标权重,进而建立信用评级方程,保证求解出的信用得分能够显著地区分违约与非违约客户。

1 最优权重的基本原理

1.1 信用评级方程中权重的重要性

信用评级方程是信用评级模型的核心,信用评级方程的好坏关系到信用评级模型的成败,而信用评级方程的权重系数直接影响着信用评级方程的评价结果,给信用评级方程不同的指标权重,

其评价结果会大相径庭,甚至截然相反.下文将给出一个例子验证上述观点.

信用评级方程中信用得分 S_i 是指标权重 w_j 的函数,如式(1)和式(2)所示.

$$S_i(\mathbf{W}) = \sum_{j=1}^m w_j x_{ij} \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, m) \quad (1)$$

$$\sum_{j=1}^m w_j = 1 \quad (2)$$

式中 $S_i(\mathbf{W})$ 是第 i 个客户的信用得分 $i = 1, 2, \dots, n$, 而 n 为客户总数; \mathbf{W} 是指标的权重向量, $\mathbf{W} = (w_1, w_2, \dots, w_j, \dots, w_m)$ w_j 表示第 j 个指标的权重; m 是信用评价指标的个数; x_{ij} 是第 i 个客户的第 j 个指标的标准化得分.

下面通过实例说明权重的极其重要性.

表 1 第 2 列和第 3 列前 3 行分别是企业 I 和企业 II 的指标数据,第 4 列和第 5 列前 3 行分别代表两组不同的权重向量(权重 I 和权重 II).

分别在权重 I 和权重 II 两种情况下计算企业 I 和企业 II 的信用得分.以企业 I 在权重 I 情况下的信用得分计算为例:将表 1 第 2 列前 3 行数据、第 4 列前 3 行数据代入公式(1)计算得企业 I 在权重 I 情况下的信用得分为 0.83,填入表 1 第 2 列第 4 行.

同理,可以得到企业 I 在权重 II 情况下的信用评分 0.24,填入表 1 第 2 列第 5 行;也可以得到企业 II 在权重 I 和权重 II 情况下的信用得分 0.31 和 0.88,分别填入表 1 第 3 列最后两行.

表 1 不同权重向量对应的信用得分

Table 1 Credit scores corresponding to different weight vectors

序号	1	2	3	4	5
	指标	企业 I	企业 II	权重 I 的 W_1	权重 II 的 W_2
1	速动比率 x_{i1}	1	0.1	0.7	0.1
2	行业景气指数 x_{i2}	0.6	0.7	0.2	0.1
3	总资产报酬率 x_{i3}	0.1	1	0.1	0.8
4	权重 I 的综合评分 S_i	0.83	0.31	—	—
5	权重 II 的综合评分 S_i	0.24	0.88	—	—

从表 1 第 2 列和第 3 列的数据可以看出,对于同一组企业的指标数据,不同的权重能够导致信用评级方程计算出完全相反的评价结果.

在权重 I 情况下,企业 I 的综合评分 0.83 大于企业 II 的综合评分 0.24;企业 I 很好,企业 II 很差;在权重 II 情况下,企业 I 的综合评分 0.31 小于企业 II 的综合评分 0.88;企业 I 很差,企业 II 很好.

两组不同的权重最终导致了完全不同的评价结果,可见权重对信用评级方程的评价结果影响极大.

综上所述,在信用评级中,不同的权重组合会导致完全不同的、甚至截然相反的评价结果.因此确定合理的权重对于建立有效的信用评级模型极为重要.

1.2 难点及解决问题的思路

难点之一是如何确定一组指标权重,使得基

于这组权重的信用评级模型,计算信用得分能够最大程度地区分违约与非违约客户.

解决的思路:在客户的信用评分与权重的函数关系中,通过以违约客户与非违约客户信用得分单位离散度的平均距离最大为目标函数,反推信用评级方程的一组最佳权重.确保信用得分能最大程度区分违约与非违约客户,改变现有研究的指标赋权与客户好坏的区分无关的弊端.

第二个难点是如何检验用本研究最优权重信用评级方程计算的信用得分对违约与非违约客户的区分能力.

解决的思路:本研究通过“查准率 (precision)”和“查全率 (recall)”刻画客户信用得分与客户违约状态之间的关系,以“查全率”为横坐标、“查准率”为纵坐标绘制 PR 曲线,以 PR 曲线的最大 F -score 作为检验信用评级方程合理性的标准.

基于违约与非违约区分程度最大化的赋权信用评级原理如图 1 所示。

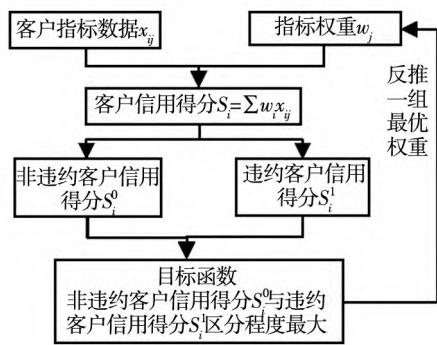


图 1 基于违约状态区分程度最大的赋权原理

Fig. 1 Weighting principle based on greatest default distinction degree

2 基于违约状态区分程度最大权重的信用评级模型

2.1 指标数据的标准化

指标数据标准化的目的是消除指标量纲影响,将指标数据转化为 $[0, 1]$ 区间的数,并将指标都转化为正向指标。

1) 定量指标标准化

① 正向指标标准化

正向指标是数值越大反映客户信用状况越好的指标,例如“速动比率”。

设 x_{ij} 是第 i 个客户第 j 个指标的标准化数据; u_{ij} 是第 i 个客户第 j 个指标的原始数据; n 是总客户数,则^[20]

$$x_{ij} = \frac{u_{ij} - \min_{1 \leq i \leq n} (u_{ij})}{\max_{1 \leq i \leq n} (u_{ij}) - \min_{1 \leq i \leq n} (u_{ij})} \quad (3)$$

② 负向指标标准化

负向指标是数值越小反映客户信用状况越好的指标,例如“资本负债率”。则^[20]

$$x_{ij} = \frac{\max_{1 \leq i \leq n} (u_{ij}) - u_{ij}}{\max_{1 \leq i \leq n} (u_{ij}) - \min_{1 \leq i \leq n} (u_{ij})} \quad (4)$$

③ 区间型指标标准化

区间型指标是数值越接近某一个理想区间时,客户信用状况越好的指标。本研究仅涉及 1 个区间型指标“居民消费价格指数”,其理想区间为 $[101, 105]$ ^[21]。

设 q_1 是理想区间的左边界; q_2 是理想区间的右边界,则 x_{ij} 为^[20]

$$x_{ij} = \begin{cases} 1 - \frac{q_1 - u_{ij}}{\max_{1 \leq i \leq n} (q_1 - \min_{1 \leq i \leq n} (u_{ij}), \max_{1 \leq i \leq n} (u_{ij}) - q_2)} & u_{ij} < q_1 \\ 1 - \frac{u_{ij} - q_2}{\max_{1 \leq i \leq n} (q_1 - \min_{1 \leq i \leq n} (u_{ij}), \max_{1 \leq i \leq n} (u_{ij}) - q_2)} & u_{ij} > q_2 \\ 1, & q_1 \leq u_{ij} \leq q_2 \end{cases} \quad (5)$$

式 (5) 的含义是原始数据 u_{ij} 在理想区间 $[q_1, q_2]$ 内时,客户信用状况最佳,标准化数据为最大值 1。原始数据在理想区间外时,原始数据 u_{ij} 距离理想区间 $[q_1, q_2]$ 越近,客户信用状况越好,标准化数据越大。

2) 定性指标标准化

对定性指标不同的状态进行打分,对反映信用状况越好的状态赋予越高的分值,从而将定性指标量化。打分标准依据已有文献^[21-22],分值均介于 0 和 1 之间,具体的打分标准如表 2 所示。

表 2 定性指标打分标准

Table 2 Standard of qualitative indicator scoring

指标	指标状态	打分
产品 销售 范围	1. 产品外销	1.00
	2. 产品内销	0.50
	3. 其他,或数据缺失	0.00
近 3 年 企业 授信 情况	1. 近 3 年有授信记录、且无违约、且无转贷	1.00
	2. 近 3 年无授信记录	0.60
	3. 近 3 年有授信记录、且无违约、但有转贷	0.50
	4. 近 3 年有违约记录、但已结清	0.30
	5. 近 3 年有违约记录,且未结清,或数据缺失	0.00

2.2 基于客户违约状态区分程度最大的最优权重确定

在信用评级方程中,每给出一组权重,就能计算出一组客户的信用得分。总有一组权重,能使违约与非违约客户的信用得分区分程度最大,这组权重就是最优权重。

2.2.1 违约与非违约客户信用得分区分程度的构建

设 D 是单位离散程度两类客户平均信用得分的距离 (standardized mean difference),用来反映违约客户与非违约客户信用得分的区分程度; n_0 是非违约客户的户数; n_1 是违约客户的户数; W 是指标的权重向量, $W = (w_1, w_2, \dots, w_j, \dots, w_m)$, 其中 w_j 表示第 j 个指标的权重; $S_i^0(W)$ 是权重向量 W 条件下第 i 个非违约客户的信用得分;

$S_i^1(\mathbf{W})$ 是权重向量 \mathbf{W} 条件下第 i 个违约客户的信用得分; $\sigma(S_i^0(\mathbf{W}))$ 是 n_0 户非违约客户信用得分的标准差; $\sigma(S_i^1(\mathbf{W}))$ 是 n_1 户违约客户信用得分的标准差. 则违约客户与非违约客户信用得分的区分程度可以表示为^[23]

$$D = \frac{\left(\sum_{i=1}^{n_0} S_i^0(\mathbf{W}) / n_0 \right) - \left(\sum_{i=1}^{n_1} S_i^1(\mathbf{W}) / n_1 \right)}{\sqrt{\sigma(S_i^0(\mathbf{W})) \sigma(S_i^1(\mathbf{W}))}} \quad (6)$$

其中

$$\sigma(S_i^0(\mathbf{W})) = \sqrt{\frac{1}{n_0} \sum_{i=1}^{n_0} \left(S_i^0(\mathbf{W}) - \sum_{i=1}^{n_0} S_i^0(\mathbf{W}) / n_0 \right)^2} \quad (7)$$

$$\sigma(S_i^1(\mathbf{W})) = \sqrt{\frac{1}{n_1} \sum_{i=1}^{n_1} \left(S_i^1(\mathbf{W}) - \sum_{i=1}^{n_1} S_i^1(\mathbf{W}) / n_1 \right)^2} \quad (8)$$

式(6)右边分子第1项 $\sum_{j=1}^{n_0} S_j^0(\mathbf{W}) / n_0$ 表示 n_0 户非违约客户信用得分的平均值; 第2项 $\sum_{j=1}^{n_1} S_j^1(\mathbf{W}) / n_1$ 表示 n_1 户违约客户信用得分的平均值; 整个分子 $\left(\sum_{i=1}^{n_0} S_i^0(\mathbf{W}) / n_0 \right) - \left(\sum_{i=1}^{n_1} S_i^1(\mathbf{W}) / n_1 \right)$ 表示两类客户信用得分之间的距离. 分子越大则表示违约客户与非违约客户信用得分距离越远;

式(6)分母中的第1项 $\sigma(S_i^0(\mathbf{W}))$ 表示 n_0 户非违约客户信用得分的标准差, 代表非违约客户信用得分的离散程度; 分母中的第2项 $\sigma(S_i^1(\mathbf{W}))$ 表示 n_1 户违约客户信用得分的标准差, 代表违约客户信用得分的离散程度; 整个分母表示非违约客户与违约客户信用得分标准差的几何平均数, 代表两类客户信用得分的离散程度. 分母越小说明同类客户信用得分越聚集.

综合上述, 式(6)中 D 的含义为单位离散程度两类客户平均信用得分的距离. 它表现为违约与非违约客户信用得分距离与离散程度的比值, 也是两类客户信用得分的区分程度. 式(6)的分子越大, 分母越小, 区分程度 D 越大, 违约客户与非违约客户的信用得分交叠越少, 信用得分越能够

区分违约与非违约客户.

式(6)中的标准差通过式(7)与式(8)求解得到, 其中式(7)表示非违约客户信用得分的标准差, 是违约客户信用得分离散程度的度量. 式(8)表示违约客户信用得分的标准差, 是非违约客户信用得分离散程度的度量.

2.2.2 构建基于区分程度最大的权重反推模型

由式(6)可知, 单位离散程度两类客户平均信用得分的距离 D 是关于信用得分 S_i 的函数. 而根据式(1)可知, S_i 是关于指标权重 w_j 的函数. 因此 D 是指标权重 w_j 的函数.

以违约与非违约客户信用得分的区分程度最大为目标函数, 以指标权重 w_j 为决策变量, 建立非线性规划模型, 如式(9)~式(11)所示. 式中各参数的含义同式(6).

$$\text{Obj: max } \frac{\left(\sum_{i=1}^{n_0} S_i^0(\mathbf{W}) / n_0 \right) - \left(\sum_{i=1}^{n_1} S_i^1(\mathbf{W}) / n_1 \right)}{\sqrt{\sigma(S_i^0(\mathbf{W})) \sigma(S_i^1(\mathbf{W}))}} \quad (9)$$

$$\text{s.t. } \sum_{j=1}^m w_j = 1 \quad (10)$$

$$w_j > 0, j = 1, 2, 3, \dots, m \quad (11)$$

式(10)和式(11)是对指标权重的约束, 使得指标权重之和为1, 且各指标的权重均大于0.

式(9)~式(11)非线性规划的特色在于通过构建信用评级方程 $S = S(\mathbf{W})$ 的指标权重 w_j 与违约与否的两类客户平均信用得分的距离 D 之间的函数关系, 以违约与否的两类客户平均信用得分距离 D 最大为目标建立非线性规划, 反推一组最优权重 \mathbf{W} . 保证了信用评级方程的评级结果, 能够最大限度地区分违约与非违约客户. 改变了现有研究^[2-5, 21, 22]的权重向量与整体违约鉴别能力无关的弊端.

2.3 信用评级方程的建立

信用评级方程如式(1)所示, 为了读者阅读方便, 在此再次给出, 并重新编号为式(12).

$$S_i(\mathbf{W}) = \sum_{j=1}^m w_j x_{ij} \quad (12)$$

将求解上文第2.2.2小节的规划模型得到的指标权重向量 \mathbf{W} 代入式(12), 即为建立的信用评级方程.

本研究采用线性方程进行评分, 是因为线性

方程在信用评级中运用广泛^[6-9]. 采用线性方程的好处在于良好的解释性(interpretability) 和较低的复杂性(complexity) ,而在精度(accuracy) 方面, 式(12) 的权重向量 W 是根据上文以两类客户平均信用评分距离最大为目标反推求得, 因此由式(12) 获得的信用评分也具有很好的精度. 综合上述, 考虑了复杂性、精度、解释性 3 方面的标准, 因此采用式(12) 的方程.

2.4 基于 PR 曲线的信用评级方程合理性检验

通过 PR 曲线(precision recall curve)^[24] 检验本研究最优权重信用评级方程计算的信用得分对违约与非违约客户的区分能力.

PR 曲线的绘制涉及到两个变量, 查准率 Precision 和查全率 Recall, 而计算这两个变量需要获得 4 个参数: TD (true default) 是实际违约被判定为违约的客户数; FD (false default) 是实际违约被判定为非违约的客户数; FN (false non-default) 是实际非违约被判定为违约的客户数; TN (true non-default) 是实际非违约被判定为违约的客户数. 用违约状态判定结果分类矩阵表示这 4 个参数的关系, 如表 3 所示.

表 3 违约状态判定结果分类矩阵

Table 3 Classification matrix of default status judgment

判定违约状态	实际违约状态		
	违约(1)	非违约(0)	合计
违约(1)	TD	FN	判定违约客户数 $TD+FN$
非违约(0)	FD	TN	判定非违约客户数 $FD+TN$
合计	实际违约客户数 $TD+FD$	实际非违约客户数 $FN+TN$	客户总数

变量查准率 Precision 和查全率 Recall 的计算如式(13) 和式(14) 所示^[24].

$$Precision = TN / (FD + TN) \quad (13)$$

$$Recall = TN / (FN + TN) \quad (14)$$

式(13) 的含义为判定正确的非违约客户数与判断为非违约客户数的比率, 用来衡量模型对非违约客户判定准确度的尺度. 当改变作为阈值的信用得分 S_i , 实际违约被判定为非违约的客户数 FD 和实际非违约被判定为非违约的客户数 TN 的数值会发生变化. FD/TN 的数值越小, 查准率的值(Precision) 越大, 模型准确判定非违约客户的能力就越强.

式(14) 的含义为判定正确的非违约客户数与实际全部非违约客户数的比率, 用来衡量模型预测非违约客户能力的尺度. 当改变作为阈值的信用得分 S_i , 实际非违约被判定为违约的客户数 FN 和实际非违约被判定为非违约的客户数 TN 的数值发生变化, 式(14) 的分母($FN+TN$) 是实际非违约客户数, 为定值, 则分子 TN 数值越大, 查全率的值(Recall) 越大, 模型成功判定非违约客户的能力就越强.

将客户信用得分从大到小排列, 依次用每一个信用得分 S_i 作为判定违约与非违约的阈值, 得分大于 S_i 的客户判定为非违约, 得分小于 S_i 的客户判定为违约. 则对于每一个信用得分 S_i 就能得到对应的一组违约状态判定分类矩阵中的参数: TD_i 、 FN_i 、 FD_i 和 TN_i . 通过式(13) 和式(14) 就能计算出对应的查准率 $Precision_i$ 和查全率 $Recall_i$, 以查准率 $Precision_i$ 为纵坐标, 查全率 $Recall_i$ 为横坐标, 每一个信用得分 S_i 对应坐标平面上的一个点($Recall_i, Precision_i$) , 连接坐标平面上所有点, 即能获得 PR 曲线.

随着查全率 Recall 增加, 查准率 Precision 降低得越慢, 即 PR 曲线越向右上方凸, 说明信用得分对违约状态的区分越准确.

令

$$F-score = 2 / [(1/Precision) + (1/Recall)] \quad (15)$$

式中 $F-score$ 表示查准率 Precision 和查全率 Recall 的调和平均数, 用来表示一个信用得分作为阈值时, 对违约与非违约的区分能力, 其中最大 $F-score$ 值可以用来判定 PR 曲线向右上方凸的程度^[25], 即信用评级方程对违约的区分能力, 即用来表示信用评级方程对违约与非违约的区分能力.

式(15) 的分母为($(1/Precision) + (1/Recall)$) 根据均值不等式原理, 得到式(16).

$$\frac{1}{precision} + \frac{1}{recall} \geq \frac{2}{\sqrt{precision \times recall}} \quad (16)$$

式(16) 是根据均值不等式原理推导出的不等式关系, 当且仅当 $Precision = Recall$ 时, 式(16) 的等号成立, 此时式(15) 的分母取得最小值, 则 $F-score$ 取得最大值, 即为信用评级方程对应的最大 $F-score$.

最大 $F-score$ 的值越接近于 1, 则说明模型的

判别能力越强^[25],信用评级方程计算的得分区分违约与非违约的能力越强.

基于 PR 曲线的信用评级方程合理性检验的特色在于,通过“查准率”和“查全率”刻画客户信用得分与客户违约状态之间的关系,以 PR 曲线的最大 *F-score* 作为检验信用评级方程合理性的标准,有别于现有研究利用 ROC 曲线检验信用评级方程合理性^[21]的方法.

采用 PR 曲线代替 ROC 曲线进行模型有效性检测,是因为本研究使用的小企业贷款数据是非平衡数据,在非平衡数据下 PR 曲线相对 ROC 曲线更加敏感^[24, 25].在进行模型对比时,用 PR 曲线能够显著反映模型判别效果的差异性.

2.5 信用等级划分

采用已经被授权的国家发明专利“基于信用等级与违约损失率匹配的信用评级系统与方法”^[26],通过信用评分与违约损失率(*loss given default*, LGD)的匹配关系,对客户的信用等级进行划分.

根据式(12)计算的信用得分 S_i 从高到低,将贷款客户进行排序,并划分 AAA、AA、…、C 共 9 个等级,等级越高对应的得分越高.通过等级得分上下限的调整,等级内贷款客户发生改变,进而等级对应的违约损失率 LGD 就随之改变.通过不断地调整信用等级,总会找到满足“信用等级越高、违约损失率越低”的最优划分结果.

2.6 有关说明

不论是信用评级研究的赋权问题还是其他问题,获取较大的样本,提高模型的预测精度都是研究问题的精要所在.样本获取及处理不属于本研究的创新点,不赘述,读者可以参考有关文献.

3 实证分析

3.1 样本选取

选取了某区域性商业银行在中国 28 个省市分支行的涵盖零售业、批发业、住宿餐饮业等 12 个非工业行业的小企业贷款数据,其中共有 1 231 笔贷款的数据,其中违约样本为 35 笔,非违约样本为 1 196 笔.

3.2 指标体系与原始数据

本研究所用的指标体系是基于现有研究构建

的指标体系,即基于偏相关分析和 Probit 回归法获得的^[21].

表 4 给出了由上述方法获得的 23 个指标及其数据,其中 b 列是准则层, c 列是指标的名称, d 列是指标类型.表 4 第 1 列~第 1 231 列,第 1 行~第 23 行是 23 个指标的原始数据,其中第 1 列~第 35 列是违约企业的 23 个指标的原始数据,第 36 列~第 1 231 列是非违约企业第 23 个指标的原始数据.表 4 第 24 行第 1 列~第 1 231 列是小企业的违约状态,第 1 232 列~第 2 462 列是违约状态对应的违约标识,其中违约企业记为 1,非违约企业记为 0.

表 4 中的左半段的 1 231 列数据,来源于文献[21].

3.3 指标数据标准化

将表 4 中第 1 列~第 1 231 列的原始数据按照 d 列的指标类型,通过上文 2.1 节中的方法进行标准化.其中定量指标的数据分别根据正向指标类型、负向指标或区间指标类型代入式(3)、式(4)或式(5),定性指标根据表 2 中的打分标准,对定性指标进行标准化.将标准化后的指标数据列入表 4 第 1 行~第 23 行,第 1 232 列~第 2 462 列.

表 4 中的右半段的 1 231 列数据,也来源于文献[21].

3.4 最优权重向量的确定

1) 违约客户信用评级方程的建立

如 2.2.1 节所述,违约企业的信用得分记为 $S_i^1(W)$,非违约企业的信用得分记为 $S_i^0(W)$.

将表 4 前 23 行第 1 232 列~第 1 266 列的 35 个违约客户的数据分别代入式(12),可以得到 35 户违约企业的信用得分关于权重向量 W 的一组函数式,如式(17-1)至式(17-35)所示.

$$\begin{cases} S_1^1(W) = 0.682w_1 + \dots + 0.182w_{11} + \dots + 0.1w_{23} & (17-1) \\ \vdots \\ S_{35}^1(W) = 1w_1 + \dots + 0.481w_{11} + \dots + 0.0w_{23} & (17-35) \end{cases}$$

2) 非违约客户信用评级方程的建立

将表 4 前 23 行第 1 267 列~第 2 462 列 1 196 户非违约企业的标准化数据分别代入式(12)中,得到非违约企业信用得分函数式,如式(18-1)至式(18-1 196)所示.

$$\begin{cases} S_1^0(W) = 0.0w_1 + \dots + 0.182w_{11} + \dots + 0.65w_{23} & (18-1) \\ \vdots \\ S_{1196}^0(W) = 0.129w_1 + \dots + 0.640w_{11} + \dots + 0.0w_{23} & (18-1196) \end{cases}$$

3) 非违约客户信用得分标准差表达式的建立

表 4 小企业信用评级指标体系及数据

Table 4 Indicator system and data of small enterprises

(a) 序号	(b) 准则层	(c) 指标	(d) 指标类型	1 231 笔借据的指标原始数据 u_{ij}					1 231 笔借据的指标标准化得分 x_{ij}						
				35 笔违约借据		1 196 笔非违约借据			35 笔违约借据		1 196 笔非违约借据				
				(1) 企业 1	... (35) 企业 35	(36) 企业 36	... (1 231) 企业 1 231	(1 232) 企业 1	... (1 266) 企业 35	(1 267) 企业 36	... (2 462) 企业 1 231				
1	C ₁ 偿债能力	X ₁ 速动比率	正向	6	...	8.798	0	...	1.142	0.682	...	1	0	...	0.129
2		X ₂ 超速动比率	正向	0.177	...	0.074	0.4	...	0.667	0.031	...	0.013	0.070	...	0.118
3	C ₂ 盈利能力	X ₃ 总资产报酬率	正向	0.02	...	1.93e-7	0.03	...	0.002	0.026	...	0	0.038	...	0.003
...	
6		X ₆ 经营活动产生的现金流量净额	正向	3.407e7	...	0	1.033e7	...	1.234e6	0.675	...	0.498	0.552	...	0.505
9	C ₅ 外部宏观条件	X ₉ 行业景气指数	正向	127.96	...	130.1	137.45	...	139.5	0.626	...	0.648	0.722	...	0.742
11		X ₁₁ 城市居民人均可支配收入	正向	9 101.35	...	19 014	9 101.35	...	24 276	0.182	...	0.481	0.182	...	0.640
12	C ₆ 企业内部非财务因素	X ₁₂ 相关行业从业年限	定性	9	...	3	8	...	10	1	...	0.4	1	...	1
14		X ₁₄ 产品销售范围	定性	数据缺失	...	其他	产品内销	...	产品外销	0	...	0	0.5	...	1
23	C ₁₀ 抵质押担保因素	X ₂₄ 抵押/质押/担保/信用	—	其他企业保证	...	数据缺失	商品房产权现房抵押	...	数据缺失	0.1	...	0	0.65	...	0
24	—	是否违约	—	违约	...	违约	非违约	...	非违约	1	...	1	0	...	0

注: 表头括号内数字表示列号.

式(7)中共涉及到两个参数或参数式.

第一个参数 n_0 是非违约企业户数, 由 3.1 节可知 $n_0 = 1 196$.

第二个参数 $S_i^0(W)$ 是非违约企业的信用得分, 来源于式(18)的表达式.

将上述参数或参数式代入式(7), 可得到非违约企业信用得分标准差的表达式, 如式(19).

$$\sigma(S_i^0(W)) = \sqrt{\frac{1}{1 196} \sum_{i=1}^{1 196} \left(S_i^0(W) - \frac{\sum_{i=1}^{1 196} S_i^0(W)}{1 196} \right)^2} \quad (19)$$

4) 违约客户信用得分标准差表达式的建立

式(8)中共涉及两个参数或参数式.

第一个参数 n_1 是违约企业户数, 由 3.1 节可知 $n_1 = 35$.

第二个参数 $S_i^1(W)$ 是违约企业的信用得分, 来源于式(17)的表达式.

将上述参数或参数式代入式(8), 可得到非违约企业信用得分标准差的表达式, 如式(20).

$$\sigma(S_i^1(W)) = \sqrt{\frac{1}{35} \sum_{i=1}^{35} \left(S_i^1(W) - \frac{\sum_{i=1}^{35} S_i^1(W)}{35} \right)^2} \quad (20)$$

5) 违约与非违约企业信用得分区分程度计算

式(9)中共涉及到 6 个参数或参数表达式.

第一个参数 n_0 是非违约企业户数, 由 3.1 节可知 $n_0 = 1 196$.

第二个参数 $S_i^0(W)$ 是非违约企业的信用得分, 来源于式(18)的表达式.

第三个参数 n_1 是违约企业户数 ,由 3.1 节可知 $n_1 = 35$.

第四个参数 $S_i^1(W)$ 是违约企业的信用得分 ,来源于式(17)的表达式.

第五个参数 $\sigma(S_i^0(W))$ 是非违约企业信用得分的标准差 ,来源于式(19)的表达式.

第六个参数 $\sigma(S_i^1(W))$ 是违约企业信用得分的标准差 ,来源于式(20)的表达式.

将上述 6 个参数或参数式代入式(9) 得到违约与非违约客户信用得分的区分度 如式(21).

$$D = \frac{\left(\sum_{i=1}^{196} S_i^0(W) / 196 \right) - \left(\sum_{i=1}^{35} S_i^1(W) / 35 \right)}{\sqrt{\sigma(S_i^0(W)) \sigma(S_i^1(W))}} \quad (21)$$

6) 构建规划模型反推最优权重

以式(21) 最大为目标函数 ,以各指标的权重之和等于 1 为约束条件 ,构建数学规划模型 ,反推出最优权重向量.即

$$\text{Obj: Max } D = \frac{\left(\sum_{j=1}^{196} S_j^0(W) / 196 \right) - \left(\sum_{j=1}^{35} S_j^1(W) / 35 \right)}{\sqrt{\sigma(S_j^0(W)) \sigma(S_j^1(W))}}$$

$$\begin{aligned} \text{s.t. } & \sum_{j=1}^m w_j = 1 \\ & w_j > 0, j=1, 2, 3, \dots, m \end{aligned}$$

求解上述数学规划模型 ,将求解得到的最优权重向量列入表 5 第 3 列 ,表 5 第 1 列和第 2 列是指标的准则层和指标层 ,与表 4 中 b 列和 c 列对应.

根据各准则层中指标的权重 ,分别将各准则层中指标权重加和 ,计算出各准则层的权重 ,列入表 5 第 4 列.

应该指出 ,表 5 中前 4 个准则层包含的指标为财务指标 ,后 6 个准则层包含的指标为非财务指标^[6, 21]. 分别将表 3 前 4 个准则层和后 6 个准则层中指标权重求和 ,计算出财务指标权重之和以及非财务指标权重之和 ,列入表 5 第 5 列.

表 5 第 6 列为对比模型各指标的权重 ,将在 3.8 节模型的对比分析中解释.

表 5 指标权重

Tabel 5 Weights of indicators

序号	1	2	3	4	5	6	
	准则层	指标层	本模型权重 w_j	准则层权重	权重之和	对比模型的权重 w_j'	
1	C_1 偿债能力	X_1 速动比率	5.141×10^{-9}	0.051	财务指标的权重之和为 0.420	0.056	
∴		X_2 超速动比率	0.051	0.263		0.049	
3	C_2 盈利能力	X_3 总资产报酬率	0.091	0.104		∴	0.061
∴		∴	∴	0.002		∴	∴
6		X_6 经营活动产生的现金流量净额	0.169			0.007	
7	C_3 营运能力	X_7 应收账款周转速度	0.104				0.135
8	C_4 成长能力	X_8 总资产增长率	0.002				0.027

9	C ₅ 企业外部宏观环境	X ₉ 行业景气指数	0.083	0.230	非财务指标的 权重之和为 0.580	0.008
∴		∴	∴	0.089		∴
11		X ₁₁ 城市居民人均可支配收入	0.147	0.199		0.014
12	C ₆ 企业内部非财务因素	X ₁₂ 相关行业从业年限	0.022	3.643×10 ⁻⁹		0.024
∴		∴	∴	3.119×10 ⁻⁹		∴
14		X ₁₄ 产品销售范围	0.067	0.063		0.035
15	C ₇ 法人代表基本情况	X ₁₅ 法人代表贷款违约记录	3.748×10 ⁻⁹			0.027
∴		∴	∴			∴
20		X ₂₀ 担任该职务时间	0.061			0.042
21	C ₈ 企业基本信用情况	X ₂₁ 近三年企业授信情况	3.643×10 ⁻⁹			0.025
22	C ₉ 企业的商业信誉	X ₂₂ 企业间合同违约次数	3.119×10 ⁻⁹		0.027	
23	C ₁₀ 抵质押担保因素	X ₂₃ 抵质押担保得分	0.063		0.024	

3.5 优化结果的初步分析结论

由表 5 不难得出如下基本结论.

1) 非财务指标在小企业信用评级中具有更重要的作用.由表 5 的第 5 列可以看出,非财务指标所占权重达到了 58%,这说明相对于财务指标,非财务指标在小企业信用评级中具有更重要的作用.

2) 对于财务指标,经营活动产生的现金流量净额在小企业信用评级中最重要的.如表 5 第 3 列第 6 行可知,经营活动产生的现金流量净额的权重最大,达到了 0.169,是财务指标中权重最大的指标.

3) 对于非财务指标,城市居民人均可支配收入在小企业信用评级中最重要的.如表 5 第 3 列第 11 行所示,城市居民人均可支配收入的权重为 0.147,是非财务指标中权重最大的指标.

4) 准则层中 C₂ 盈利能力对小企业信用评级最重要的.如表 5 第 4 列所示,按权重从大到小对准则层进行排序:C₂盈利能力>C₅外部宏观条件>C₇法人代表基本情况>C₃营运能力>C₆企业内部非财务因素>C₁₀抵质押担保因素>C₁偿债能力>C₄成长能力>C₉企业基本信用情况>C₉企业的商业信誉.其中 C₂盈利能力的权重最大,为 0.263,C₉企业

的商业信誉权重最小,为 3.12×10⁻⁹.

3.6 信用评级方程的建立

将表 5 第 3 列的 w_j代入式(12),得到式(22)

$$S_i(W) = 5.141 \times 10^{-9} x_{i1} + 0.051 x_{i2} + 0.091 x_{i3} + \dots + 0.169 x_{i6} + 0.104 x_{i7} + 0.002 x_{i8} + 0.083 x_{i9} + \dots + 0.147 x_{i11} + 0.022 x_{i12} + \dots + 0.067 x_{i14} + 3.748 \times 10^{-9} x_{i15} + \dots + 0.061 x_{i20} + 3.643 \times 10^{-9} x_{i21} + 3.119 \times 10^{-9} x_{i22} + 0.063 x_{i23} \quad (22)$$

式(22)即为本研究建立的小企业信用评级方程.

将表 4 前 23 行第 1 232 列~第 2 462 列的标准化数据代入式(22),得到 1 231 户贷款企业的信用得分 S_i,列入表 6 第 3 列第 1 行~第 1 231 行.

表 6 前 1 231 行第 1 列是企业编号;第 2 列是违约标识,违约标识为 1,非违约标识为 0;第 4 列是对比模型计算的信用得分 S_i',将在 3.8 节中解释;第 5 列和第 6 列分别为年应收本息和年未收本息,将在 3.9 节信用等级划分中使用.表 6 第 1 232 行第 3 列和第 4 列分别是本模型和对比模型 PR 曲线的最大 F-score 值,将在 3.7 节和 3.8 节中解释.

表 6 小企业信用得分及违约状态

Tabel 6 Credit scores and default status of small enterprises

序号	1	2	3	4	5	6
	企业编号	违约标识	本模型信用得分 S _i	对比模型信用得分 S _i '	年应收本息	年未收本息
1	企业 1	1	0.227	0.100	13 066 243.93	3 196.75
∴	∴	∴	∴	∴	∴	∴
35	企业 35	1	0.310	0.249	115 097 967.47	65 248.66

36	企业 36	0	0.400	0.283	5 385 896.88	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1 231	企业 1 231	0	0.473	0.354	20 600 154.57	0
1 232	PR 曲线检验的最大 F -score 值		0.991	0.980	—	—

3.7 信用评级方程的合理性检验

本研究用 PR 曲线方法对信用评级方程的合理性进行检验.

将表 6 第 3 列前 1 231 行小企业信用得分数据按从大到小顺序重新排列,如表 7 第 2 列所示.表 7 第 3 列为违约标识,它与企业的对应关系与表 4 相同.

从表 7 第 1 行开始,分别以每一个得分 S_i 作为阈值,得分大于 S_i 的企业判定为非违约,得分小于 S_i 的企业判定为违约.根据企业的判定违约状态和实际违约状态,计算得分 S_i 对应的查准率 $Precision_i$ 和查全率 $Recall_i$.

以排序后第 50 位的企业得分 $S_{50} = 0.514$ 作为阈值计算查准率 $Precision_{50}$ 和查全率 $Recall_{50}$ 测算实例,其他类推.

信用得分大于或等于 0.514 的企业判定为非违约企业,信用得分小于 0.514 的企业判定为违约企业.

根据判定的违约状态和实际的违约状态可以按照表 3 的分类方法将 1 231 户企业分为 4 类,其中,实际违约被判定为违约的企业数 TD_{50} 为 35 户,列入表 7 第 50 行第 4 列;实际违约被判定为非违约的企业数 FD_{50} 为 0 户,列入表 7 第 50 行第 5 列;实际非违约被判定为违约的企业数 FN_{50} 为 1 146 户,列入表 7 第 50 行第 6 列;实际非违约被判定为非违约的企业数 TN_{50} 为 50 户,列入表 7 第 50 行第 7 列.

将表 7 第 50 行第 5 列的数据 $FD_{50} = 0$ 和第

50 行第 8 列的数据 $TN_{50} = 50$ 代入式(13),算出第 50 个得分作为阈值的查准率 $Precision_{50} = TN_{50} / (FD_{50} + TN_{50}) = 1$,列入表 7 第 50 行第 8 列.

将表 7 第 50 行第 4 列 $TD_{50} = 35$ 和第 50 行第 6 列 $FN_{50} = 1 146$ 代入式(14)可以计算出查全率 $Recall_{50} = TN_{50} / (TD_{50} + FN_{50}) = 0.042$,列入表 7 第 50 行第 9 列.

以查全率 $Recall_{50}$ 为横坐标,查准率 $Precision_{50}$ 作为纵坐标,则可在平面上获得 1 个点 $(0.042, 1)$.

同理,以每一个信用得分 S_i 作为阈值,都能获得一组查准率 $Precision_i$ 和查全率 $Recall_i$,如表 7 第 8 列和第 9 列.

以表 7 第 9 列查全率 $Recall_i$ 为横坐标,以表 7 第 8 列查准率 $Precision_i$ 为纵坐标,绘制本模型信用得分对应的 PR 曲线,如图 2 中实线所示.

计算 PR 曲线的最大 F -score 值.根据上文 2.4 节中的证明,当 $Precision_i$ 和 $Recall_i$ 相等时, F -score 取得最大值.根据表 7 第 1 196 行,当信用得分为 $S_{1196} = 0.261$ 时, $Precision_i$ 和 $Recall_i$ 相等.

将表 7 第 1 196 行第 8 列~第 9 列的 $Precision_{1196} = 0.991$ 和 $Recall_{1196} = 0.991$ 代入式(15),计算出本模型 PR 曲线的最大 F -score 值为 0.991,填入表 6 第 1 232 行第 3 列.

本模型的最大 F -score 值非常接近 1,说明本模型能通过 PR 曲线的检验,具有很强的违约鉴别能力^[24].

表 7 按从大到小排序的企业信用得分为阈值计算查准率和查全率

Table 7 Calculate the precision and recall through enterprise credit points sorted by descending order

序号	1	2	3	4	5	6	7	8	9
	企业编号	信用得分 S_i	违约标识	TD_i	FD_i	FN_i	TN_i	$Precision_i$	$Recall_i$
1	970	0.585	0	35	0	1 195	1	1	0.000 84
2	670	0.568	0	35	0	1 194	2	1	0.001 67
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
50	926	0.514	0	35	0	1 146	50	1	0.042 30
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1 196	6	0.261	1	24	11	11	1 185	0.991	0.991

∴	∴	∴	∴	∴	∴	∴	∴	∴	∴
1 230	19	0.126	1	1	34	0	1 196	0.972	1
1 231	20	0.126	1	0	35	0	1 196	0.972	1

3.8 信用评级模型对比分析

本研究构建的信用评级模型作为模型 1.

基于同一指标体系、采用变异系数进行赋权的信用评级模型^[21]作为模型 2.

由于模型 1 和模型 2 采用的是同一个指标体系,故可比,也就是说通过模型 1 与模型 2 进行对比能够反映本研究所提出的赋权方法的优劣.

模型 2 的权重如表 5 第 6 列所示^[21].将表 5 第 6 列的数据以及表 4 第 1 232 列~第 2 462 列共计 1 231 列的数据代入式(12)可以算出模型 2 (对比模型)的企业信用得分 S_i ,列入表 6 第 4 列第 1 行~第 1 231 行.

模型 1 与模型 2 的比较.

基于表 6 第 1 行~第 1 231 行第 2 列和第 4 列的数据,依照 3.7 节的方法,在图 2 中用虚线画出模型 2 的 PR 曲线^[24],实线是模型 1 对应的 PR 曲线^[24].

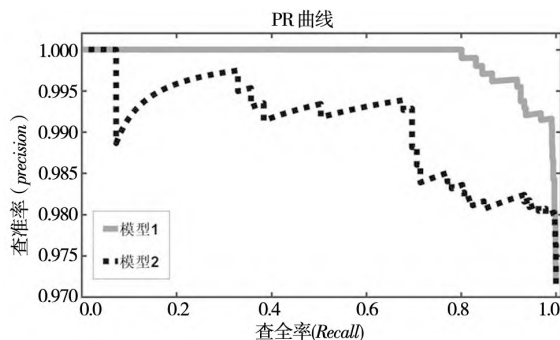


图 2 模型 1 与模型 2 的 PR 曲线比较

Fig.2 PR curves corresponded to two credit scoring models

模型 1 的 PR 曲线(实线)几乎完全在模型 2 的 PR 曲线(虚线)上方,模型 1 的 PR 曲线向右上凸的程度明显大于模型 2,也即使用模型 1 能够在查全率逐渐增加时,一直保持极高的查准率.

同 3.7 节,计算模型 2 对应 PR 曲线的最大 F -score 值^[25]为 0.980,列入表 6 第 1 232 行第 4 列.通过最大 F -score 值对比可知,模型 1 (最大 F -score 值为 0.991) 相对于模型 2 具有更高的最大 F -score 值,也就是说模型 1 具有更好的违约鉴别能力,所以本研究提出的赋权方法具有明显优势.

3.9 信用等级划分

本研究通过国家授权发明专利^[26 27]进行信用等级划分.

按照表 6 第 3 列的本模型计算的信用评分 S_i 从高到低,将贷款企业重新进行排序,并初步划分为 9 个等级,根据表 6 第 5 列和第 6 列的数据分别算出每一个等级的违约损失率 LGD^[26 27].通过调整等级得分上下限,每个等级对应的违约损失率 LGD 随之改变,总会找到满足“信用等级越高、违约损失率越低”的最优划分结果.

划分结果如图 3 所示.图中每个等级对应的横轴长度表示这个等级对应的违约损失率大小.

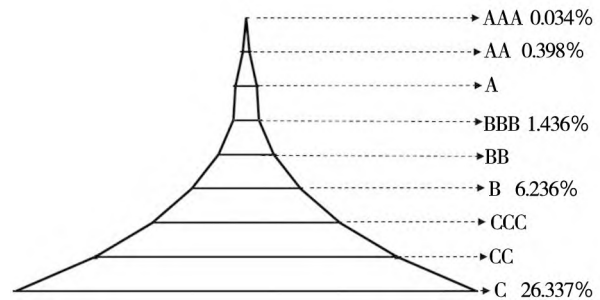


图 3 信用等级对应的违约损失率

Fig.3 LGD corresponded to nine credit ratings

由图 3 可知,本研究的信用等级划分结果满足“信用等级越高、违约损失率越低”的评级本质规律.

4 结束语

4.1 主要结论

1) 对于小企业信用评级,非财务因素比财务因素具有更重要的作用;对于财务指标,“经营活动产生的现金流量净额”指标在小企业信用评级中最重要;对于非财务指标,“城市居民人均可支配收入”指标在小企业信用评级中最重要.在所有 10 个准则层中 C_2 盈利能力对小企业信用评级的最重要.

2) 本研究基于客户违约状态区分程度最大得到的信用评级方程,相对于现有通过单个指标违约鉴别力进行赋权的信用评级方程,具有更高的违约鉴别能力.由本研究构建的信用评级方程

对应的最大 F -score 值为 0.991, 高于通过变异系数赋权的评级方程的最大 F -score 值(0.980) [21].

4.2 主要创新

1) 通过构建信用评级方程 $S=S(W)$ 的权重向量 W 与违约与否的两类客户单位离散度的平均信用得分距离 D 之间的函数关系, 以违约与否的两类客户单位离散度平均信用得分距离 D 最大为目标建立非线性规划, 反推一组最优权重 W , 保证了信用评级方

程的评级结果, 能够最大限度地区分违约与非违约客户, 改变了现有研究^[2-5, 21, 22]的权重向量与整体违约鉴别能力无关的弊端.

2) 根据 PR 曲线最大 F -score 值来检验信用评级方程的有效性, 最大 F -score 值越大, 信用评级方程越能显著区分违约与非违约企业, 确保信用评级方程构建的合理性.

参考文献:

- [1] Benmelech E, Dlugosz J. The credit rating crisis [J]. NBER Macroeconomics Annual, 2010, 24(1): 161-208.
- [2] Che Z H, Wang H S, Chuang C L. A fuzzy AHP and DEA approach for making bank loan decisions for small and medium enterprises in Taiwan [J]. Expert Systems with Applications, 2010, 37(10): 7189-7199.
- [3] 张目, 黄春燕, 李岩. 基于相对熵和可变模糊集理论的战略新兴产业企业信用评价 [J]. 数学的实践与认识, 2014, 44(13): 18-26.
Zhang Mu, Huang Chunyan, Li Yan. Credit evaluation of enterprises in strategic emerging industry based on relative entropy and variable fuzzy sets theory [J]. Mathematics in Practice and Theory, 2014, 44(13): 18-26. (in Chinese)
- [4] 陈晓红, 杨志慧. 基于改进模糊综合评价法的信用评估体系研究——以我国中小上市公司为样本的实证研究 [J]. 中国管理科学, 2015, 23(1): 146-153.
Chen Xiaohong, Yang Zhihui. The study of credit evaluation system based on improved fuzzy evaluation method: Evidence from small and medium-sized enterprises in China [J]. Chinese Journal of Management Science, 2015, 23(1): 146-153. (in Chinese)
- [5] 李步军, 高玉红, 王继顺. 基于灰区间关联分析的银行个人信用优化评估模型 [J]. 数学的实践与认识, 2016, 46(9): 289-292.
Li Bujun, Gao Yuhong, Wang Jishun. The optimized personal credit assessment model of the bank based on grey interval relational analysis [J]. Mathematics in Practice and Theory, 2016, 46(9): 289-292. (in Chinese)
- [6] Altman E I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy [J]. The Journal of Finance, 1968, 23(4): 589-609.
- [7] Desai V S, Crook J N, Overstreet G A. A comparison of neural networks and linear scoring models in the credit union environment [J]. European Journal of Operational Research, 1996, 95(1): 24-37.
- [8] 王春峰, 李文华. 商业银行信用风险评估: 投影寻踪判别分析模型 [J]. 管理工程学报, 2000, 14(2): 43-46.
Wang Chunfeng, Li Wenhua. Credit risk assessment in commercial banks' projection pursuit discriminant model [J]. Journal of Industrial Engineering/Engineering Management, 2000, 14(2): 43-46. (in Chinese)
- [9] Grablowsky B J, Talley W K. Probit and discriminant functions for classifying credit applicants—a comparison [J]. Journal of Economics and Business, 1981, 33(3): 254-261.
- [10] Duca J V, Whitesell W C. Credit cards and money demand: A cross-sectional study [J]. Journal of Money, Credit and Banking, 1995, 27(2): 604-623.
- [11] Hand D J, Henley W E. Statistical classification methods in consumer credit scoring: A review [J]. Journal of the Royal Statistical Society: Series A (Statistics in Society), 1997, 160(3): 523-541.
- [12] 李战江. 基于分层逻辑回归的小企业信用评价模型 [J]. 统计与决策, 2016, (7): 178-182.
Li Zhanjiang. Small business credit evaluation model based on hierarchical logic regression [J]. Statistics and Decision, 2016, (7): 178-182. (in Chinese)
- [13] 程砚秋. 基于不均衡数据的小企业信用风险评价 [J]. 运筹与管理, 2016, 25(6): 181-189.
Cheng Yanqiu. Credit rating of small enterprises based on unbalanced data [J]. Operations Research and Management Science, 2016, 25(6): 181-189. (in Chinese)

- [14] West D. Neural network credit scoring models [J]. *Computers & Operations Research*, 2000, 27(11): 1131–1152.
- [15] 庞素琳. 概率神经网络信用评价模型及预警研究 [J]. *系统工程理论与实践*, 2005, 25(5): 43–48.
Pang Sulin. Study on credit scoring model and forecasting based on probabilistic neural network [J]. *Systems Engineering: Theory & Practice*, 2005, 25(5): 43–48. (in Chinese)
- [16] Tsai C F, Wu J W. Using neural network ensembles for bankruptcy prediction and credit scoring [J]. *Expert Systems with Applications*, 2008, 34(4): 2639–2649.
- [17] 王春峰, 康 莉. 基于遗传规划方法的商业银行信用风险评估模型 [J]. *系统工程理论与实践*, 2001, 21(2): 73–79.
Wang Chunfeng, Kang Li. A model based on genetic programming for credit risk assessment in commercial banks [J]. *Systems Engineering: Theory & Practice*, 2001, 21(2): 73–79. (in Chinese)
- [18] Ong C S, Huang J J, Tzeng G H. Building credit scoring models using genetic programming [J]. *Expert Systems with Applications*, 2005, 29(1): 41–47.
- [19] Abdou H A. Genetic programming for credit scoring: The case of Egyptian public sector banks [J]. *Expert Systems with Applications*, 2009, 36(9): 11402–11417.
- [20] 迟国泰, 王 卫. 基于科学发展的综合评价理论、方法与应用 [M]. 北京: 科学出版社, 2009.
Chi Guotai, Wang Wei. *Comprehensive Evaluation Theory, Method and Application Based on the Scientific Development* [M]. Beijing: Science Press. 2009. (in Chinese)
- [21] 迟国泰, 张亚京, 石宝峰. 基于 Probit 回归的小企业债信用评级模型及实证 [J]. *管理科学学报*, 2016, 19(6): 136–156.
Chi Guotai, Zhang Yajing, Shi Baofeng. The debt rating for small enterprises based on Probit regression [J]. *Journal of Management Sciences in China*, 2016, 19(6): 136–156. (in Chinese)
- [22] 赵志冲, 迟国泰. 基于似然比检验的工业小企业债信用评级研究 [J]. *中国管理科学*, 2017, 25(1): 45–56.
Zhao Zhichong, Chi Guotai. Facility rating of small industrial enterprises based on likelihood ratio test [J]. *Chinese Journal of Management Science*, 2017, 25(1): 45–56. (in Chinese)
- [23] Rezac M, Rezac F. How to measure the quality of credit scoring models [J]. *Czech Journal of Economics and Finance (Finance a uver)*, 2011, 61(5): 486–507.
- [24] Buckland M, Gey F. The relationship between recall and precision [J]. *Journal of the American Society for Information Science*, 1994, 45(1): 12.
- [25] Goutte C, Gaussier E. A Probabilistic Interpretation of Precision, Recall and F-score, with Implication for Evaluation [C]// *ECIR'05 Proceedings of the 27th European conference on Advances in Information Retrieval Research*. 2005: 345–359.
- [26] 迟国泰, 石宝峰. 基于信用等级与违约损失率匹配的信用评级系统与方法 [P]. 中国: ZL201210201461.6. 2015–08–19.
Chi Guotai, Shi Baofeng. Credit Rating System and Method Based on Matching of Credit Rating and LGD [P]. China: ZL201210201461.6. 2015–08–19. (in Chinese)
- [27] 迟国泰, 程砚秋. 基于信用等级与违约损失率匹配的信用评级调整方法 [P]. 中国: ZL 2012 1 0201114.3. 2015–11–18.
Chi Guotai, Cheng Yanqiu. Credit rating adjustment method based on Matching of Credit Rating and LGD [P]. China: ZL 201210201114.3. 2015–11–18. (in Chinese)

Credit rating model based on weight of greatest default distinction degree

ZHOU Ying

School of Economics & Management, Dalian University of Technology, Dalian 116024, China

Abstract: Credit scoring model is the core of credit rating model, and the weight is the key factor which influ-

ences the accuracy of credit scoring. A rational weight vector is of crucial importance to credit rating. This paper establishes a nonlinear programming which aims at maximizing a standardized mean difference of the credit scores of two kinds of clients. Then by solving the programming, the optimal weight of credit scoring model is derived. To guarantee the effectiveness, the model is validated by Precision Recall Curve. Next, altogether 1231 enterprises are rated based on the credit score calculated by our model. The innovations and characters in this paper are: firstly, an objective function is established by constructing the function relationship between the weight vector W and the standardized mean difference D of the two group of clients. Secondly, a nonlinear programming model, which maximizes the above mentioned objective function, is established to find the optimal weight vector W . This method can guarantee that the scores based on the weights have a significant default discrimination power and can distinguish the default to the largest possible extent, thus avoiding the drawbacks of recent researches. Secondly, non-financial indexes, accounting for nearly 58%, are more important than financial indexes on small enterprise credit rating. Among the non-financial indexes, per capital disposable income of urban residents is most important, with a weighting of 14.7%. Thirdly, the credit scoring model is validated by PR curve method. The research result shows that our credit scoring model has a better discrimination power than classic models in existing researches.

Key words: credit rating; weight vector; optimal weighting; small enterprise credit rating