

信用评分模型的设计与决策分析^①

詹原瑞 田宏伟

(天津大学管理学院 300072)

摘要 本文对商业银行信贷业务中的贷款者的信用问题,运用决策分析提出了信用评分的设计原理及临界分值的确定方法,给出了最优贷款策略的期望收益和风险的计算公式。最后,以一例说明之。

关键词 信用风险 可能机会 机会影响图 决策分析

1 引言

当前在社会主义市场经济运行的环境中,我国商业银行面临的主要风险是信用风险,其突出的表现是不良贷款的比例偏高,较大数量的贷款(其中大部分是历年累计的无担保信用贷款)不能到期回收本息^[1]。商业银行不良贷款的比例偏高首先表现为我国银行主力的国有独资商业银行上。这对我国的经济运行和社会安定具有极大的威胁^[2]。因此,在今后一个时期商业银行不仅要确立正确的防范金融风险的指导思想,还应建立科学的贷款决策体系,其核心是贷款决策的科学化。在贷款决策中除了定性分析之外,还要大力引入定量分析。其中贷款者的信用分析是防范信用风险的关键所在。商业银行信贷经营的成败在很大程度上取决信用分析是否科学可靠。

本文拟就我国商业银行贷款中的一项崭新业务—消费者贷款业务,用决策分析的方法设计信用评分模型。消费者贷款是银行或非银行的金融机构向以个人或家庭为对象发放用于购买消费品或支付其他费用的贷款。这种贷款是目前发达国家商业银行的重要信贷业务,在我国这项业务方兴未艾^[3]。现已有多家银行开办住房贷款、购车贷款和信用卡等项业务。信用评分是指个人向银行提出贷款申请时,银行对申请者能否按期还本付息的信用状况进行评估。银行根据结果,即申请者按期还款的可能性的预测,决定是否接受申请者的贷款要求。

下面我们在第二段用机会影响图讨论贷款信用问题及信用评分模型的设计,第三段用决策树分析银行最优信贷策略得出信用记分临界值,最后在第四段用实例说明上述模型的建立与应用及注意事项。

2 信用评分模型的设计

在考虑消费者贷款信用时,提出了申请者按时还款可能性的预测问题。银行在作预测申请者信用时可利用申请表中的信息,一般申请表中含有年龄、职业、收入、现职工作年限、住房是个人的还是租借、银行帐户的类型和数量及持有的主要信用卡等。这些项的内容都可看作随机变量,其结局是评估申请者信用的有用信息。我们对贷款申请者的各种因素的评价用还款的可能机会来度量的。假如年收入2~2.5万元的贷款申请者还款的可能性是95%,则银行收回贷款的机

^①本文1998年6月24日收到。

会是 19/1，即银行收回该人贷款与收不回贷款是 19:1。采用机会评估信用的原因是便于利用历史数据。例如，现有数据表明在以往的 40 笔贷款中有 38 笔收回贷款，2 笔变成了呆帐，则可预测一笔同类贷款收回的机会是 38/2。为了简化起见，下面只考虑收入和住房两个因素且假设二者相互独立以此来说明申请者信用的先验机会与后验机会间的关系。

机会影响图是表示不确定变量间关系的最有用的工具，是所有结点都是表示随机变量的无环有向图^[4]。其中结点间的有向弧表示两个变量是相关的；否则表示两个变量是无条件或有条件独立。图 1 是表示这个简化的贷款申请者信用的机会影响图。图 1 (a) 表示根据历史数据估计申请者信用好或差的边际概率 $p_X(G)$ 和 $p_X(B)$ ，确定已知申请者信用价值 (X)，预测收入 (Y) 和住房 (Z) 的似然率 $p_{Y|X}(y|x)$ 和 $p_{Z|X}(z|x)$ 。我们的目标是翻转图 1 (a) 中的两条有向弧的方向，变成已知 $Y=y, Z=z$ 条件下，预测申请者信用价值如图 1 (b) 所示。根据贝叶斯定理利用图 1 (a) 所示的预测似然率计算决策概率 $p_{X|Y,Z}(G|y,z)$ 与 $p_{X|Y,Z}(B|y,z)$ 。

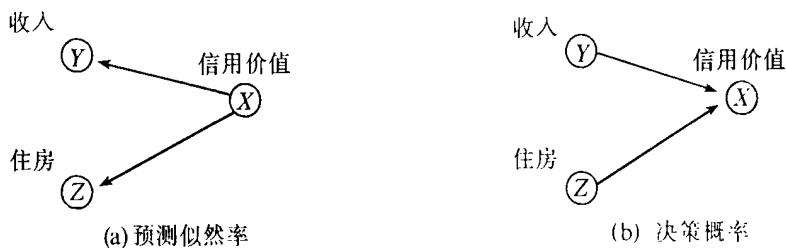


图 1 贷款申请者信用问题的机会影响图

在决策问题中定量表示不确定性一般用主观概率，此外还有表示不确定事件发生的“机会”。机会可用事件发生与不发生的概率之比定义，譬如申请者信用好的先验机会是申请者信用好与差的先验概率的比率

$$O_X(G) = \frac{p_X(G)}{p_X(B)} = \frac{p_X(G)}{1 - p_X(G)} \tag{1}$$

若贷款申请者还款的可能性是 95%，则银行收回贷款的机会是 19/1，即 $(0.95 / (1 - 0.95))$ 。根据贝叶斯定理可得，已知 Y 和 Z 取值，计算 X 的后验机会公式

$$O_{X|Y,Z}(G|y,z) = \frac{O_{X|Y}(G|y) O_{X|Z}(G|z)}{O_X(G)}$$

为了适合实际应用的需要，将上式的乘除运算简化为加减运算，在上式两边取自然对数 \ln 得

$$\ln(O_{X|Y,Z}(G|y,z)) = \ln(O_{X|Y}(G|y)) + \ln(O_{X|Z}(G|z)) - \ln(O_X(G)) \tag{2}$$

式 (2) 是根据只考虑两个因素收入和住房的分析得到的结果，若考虑多项因素只需在式 (2) 增加求和的项数。

因为事件发生机会的对数一般数值较小又不是整数，用这样数作申请者信用的记分不方便，所以评分模型的记分一般采用将上面的数值经过线性变换得到较大的整数值。例如一个人信用好的机会为 $O(G) = 20$ ， $\ln(O(G))$ 值 s 为 2.996，在表 1 中的整数得分 s' 是 200。

2.1 记分与预测的记法

贷款申请者的实际信用价值用随机变量 X 表示，其结局 $X = \{G, B\}$ ，其中 G 和 B 分别代表信用好与差，具体含义视决策问题而定。例如，不欺骗与欺骗，不倒闭与倒闭，三个月还款与不

还款等。根据申请表中的内容确定申请者信用价值的得分，用随机变量 S 表示，其结局

$$S = \{s_1, s_2, \dots, s_n\}。$$

银行的决策 D 是接受 A 还是拒绝 R 贷款者的申请， $D = \{A, R\}$ 。决策结果 $V(X, D)$ 取决于是否贷款 D 和申请人的信用价值 X ，其可能的取值集合是 $V = \{0, g, -l\}$ ，其中 g 是收益， l 是损失(恒为正数)，即 $V(G, A) = g$ ， $V(B, A) = -l$ 及 $V(G, R) = V(B, R) = 0$ 。信用得分 s 的边际概率为 $p_S(s)$ ，申请者信用价值的先验概率 $p_X(G)$ 与 $p_X(B)$ 是不随记分系统改变而改变。

2.2 记分变换

根据上述分析知，贷款申请者的信用得分 s 是申请者信用好的机会的自然对数值，得分 s 的申请者信用好的概率预测值为 $f(s)$ 。二者间有如下关系式

$$s = \ln \frac{f(s)}{1 - f(s)} \tag{3}$$

和

$$f(s) = \frac{1}{1 + e^{-s}} \tag{4}$$

由此可见 s 和 $f(s)$ 是等价的，二者可互换。

3 决策分析

银行根据上述信用评分模型得到贷款申请者信用得分，依此决定是否贷款的决策问题如图 2 的决策树所示^[5]。在图 2 中的 n 个决策小树中，只有 4 个小树标上数据。在 F 结点引出的分枝

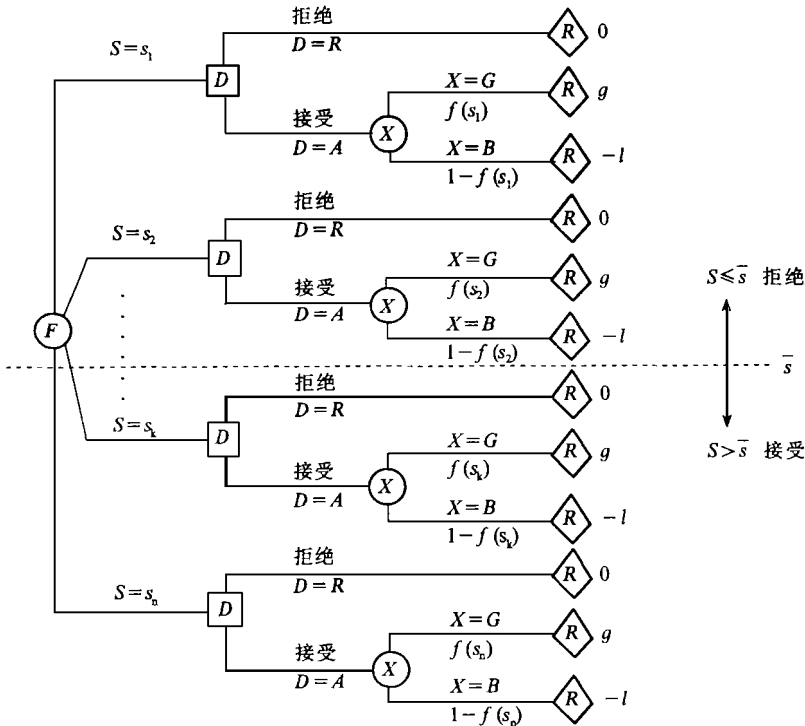


图 2 信贷决策问题的决策树

上以得分 s_i 给出的预测结果。对每个 s_i , 申请者信用好的概率是 $f(s_i)$, 信用差的概率是 $1-f(s_i)$, 并且 s_i 按递增顺序由上至下排列。这时, 信贷决策变成确定接受贷款请求的最低信用得分, 即临界值 \bar{s} 。对金融机构而言, 一般都是主观地决定单一的临界记分值 \bar{s} , 拒绝得分等于或小于临界得分的申请者, 接受得高于临界值的申请。在这里, 临界记分值 \bar{s} 的确定是以银行由信用得分 s 的贷款获得的期望收益应等于银行无风险投资的收益为依据的, 无风险投资的收益如购买同期国债的利息收入。设无风险利率为 r , 由图 2 知当信用得分为临界值 \bar{s} 的期望收益为

$$\begin{aligned} U(\bar{s}) &= g f(\bar{s}) - l(1-f(\bar{s})) = rl \\ f(\bar{s}) &= \frac{(1+r)l}{g+l} \end{aligned} \quad (5)$$

式中 $f(\bar{s})$ 也称信贷决策的无差概率。

由式(5)计算的临界值 \bar{s} 便确定了最优的信贷策略如图 2 所示:

当申请者的信用得分 s 等于或小于临界得分 \bar{s} 时, 银行拒绝其贷款申请;

当申请者的信用得分 s 大于临界得分 \bar{s} 时, 银行接受其贷款申请。

3.1 最优信贷策略的期望收益与风险

当银行采用上述最优信贷策略, 由一个信用得分 s 贷款申请者获得的期望收益是

$$U(s)^* = E_{X|S}[V(X, d^*) | S = s] \\ U(s)^* = \begin{cases} 0 & s \leq \bar{s} \\ gf(s) - l(1-f(\bar{s})) & s > \bar{s} \end{cases} \quad (6)$$

将式(5)代入(6)得

$$U(s)^* = \begin{cases} 0 & s \leq \bar{s} \\ (g+l)(f(s)-f(\bar{s})) + rl & s > \bar{s} \end{cases} \quad (7)$$

因为 $f(s) > f(\bar{s})$, 故 $v(s)^*$ 总为正。

此时, 若以收益的标准差表示贷款风险, 则银行的信用风险是

$$\sigma_s^* = \begin{cases} 0 & s \leq \bar{s} \\ (g+l) \sqrt{f(s)(1-f(s))} & s > \bar{s} \end{cases} \quad (8)$$

由式(7)和(8)可知: 若已知 \bar{s} 固定, 信用得分越高的申请者 ($f(s)$ 非常接近 1) 的贷款使银行获得的期望收益越大, 承担的风险越小。但是, 若银行提高信用记分的临界值 \bar{s} , 银行在降低风险的同时期望收益也减少。因此银行确定 \bar{s} 取决 g 、 l 和 r 的正确地选择。

3.2 最优的无条件期望收益与风险和总期望收益与风险

上面我们已讨论银行对以信用记分 s 贷款申请者拒绝或接受问题。然而贷款组合 (即那些被接受的申请者) 的无条件期望利润取决: (1) 在研究总体中, 每个人得分 s 的概率边际分布, $p_s(s) = Pr(S=s)$, (2) 已知申请者的信用得分 s , 他是信用好的概率是 $f(s) = p_{X|S}(G|s)$, (3) 银行单位贷款的增益 g 、损失 l 和无风险利率 r , 即接受/拒绝最优策略方针 \bar{s} 。

假设银行只向信用得分 $s > \bar{s}$ 申请者贷款, 并假设记分独立于增益和损失, 并保证从接受每个申请贷款都得到正的期望利润, 以致整个组合的期望利润保证为正。如果将根据样本得到的数据应用到总体中, 对式(7)中期望收益用记分的边际概率 $p_s(s)$ 加权, 得到无条件最优期望收益和风险分别是

$$\begin{aligned}
 U^* &= \sum_S E_{X|S} [R(X; d^*) | S] p_s(s) \\
 &= (1+g) \left[\sum_{S>\bar{S}} f(s) p_s(s) - f(\bar{S}) \sum_{S>\bar{S}} p_s(s) \right] + r l \sum_{S>\bar{S}} P_s(s) \tag{9}
 \end{aligned}$$

$$\sigma^* = \sqrt{\sum_{S>\bar{S}} P_s^2(s) \sigma_s^{*2}} \tag{10}$$

当考虑所有的申请者时, 银行从此项业务获得的总期望收益与风险分别是

$$V = \sum_{S>\bar{S}} n_s U(s)^* = n U^* \tag{11}$$

其中 n 是贷款申请总人数, n_s 是信用得分为 $S(S > \bar{S})$ 人数。

$$\sigma = \sqrt{n} \sigma^* = \sqrt{n \sum_{S>\bar{S}} P_s^2(s) \sigma_s^{*2}} \tag{12}$$

式 (11) 和 (12) 说明银行开展信贷营销的总期望收益是随着所占市场份额而线性增加的, 但所冒的风险是按平方根增加, 即相对减少。所以银行致力于开拓新兴市场是增加收益减少风险的重要途径。但须注意开发市场的成本不能超过一定的界限。

4 举例

现以一家外国商业银行的信贷业务为例说明上述模型的应用^[6]。根据历史数据得贷款申请者得分分布如表 1 所示。当设计出信用评分模型并获得了估计值, 习惯上, 用申请者的信用得分识别在给定总体中的每个申请者的信用价值。再将申请者划成如表 1 中第 1 列得分区间某一组内, 第 2 和 3 列给出样本内信用得分在某一区间内, 申请者是信用好或差的人数, 第 5 列给出信用得分是第一列所示区间中申请者是信用好的机会对数 s , 例如, 考虑一申请者信用得分 s' 在 180 到

表 1 贷款信息问题的数据

申请者信用得分(s')	样本数据			信用好对数的记分(s)	预测似然率		决策概率		记分分布 $p_s(s)$	万元	
	信用好	信用差	总数		$p_{S X}(s G)$	$p_{S X}(s B)$	$p_{X S}(G s) = f(s)$	$p_{X S}(B s) = 1 - f(s)$		$r(s)^*$	σ_s^*
低于 170	1386	281	1667	1.60	0.147	0.493	0.831	0.169	0.166	0.000	0.000
170- 179	760	70	830	2.38	0.080	0.123	0.916	0.084	0.083	0.000	0.000
180- 189	851	63	914	2.60	0.090	0.111	0.931	0.069	0.091	0.000	0.000
190- 199	940	48	988	2.97	0.099	0.084	0.951	0.049	0.099	0.000	0.000
200- 209	905	39	944	3.14	0.096	0.069	0.959	0.041	0.094	0.103	0.228
210- 219	920	25	945	3.61	0.097	0.044	0.974	0.026	0.094	0.120	0.183
220- 229	931	17	948	4.00	0.099	0.030	0.982	0.018	0.095	0.129	0.153
230- 239	857	12	869	4.27	0.091	0.021	0.986	0.014	0.087	0.134	0.135
240- 249	770	7	777	4.70	0.081	0.012	0.991	0.009	0.078	0.140	0.109
250- 259	586	5	591	4.76	0.062	0.009	0.992	0.008	0.059	0.140	0.102
260- 289	433	1	434	6.07	0.046	0.002	0.998	0.002	0.043	0.148	0.051
290和以上	111	1	112	-	0.012	0.002	0.991	0.009	0.011	0.140	0.109
合计	9450	569	10,019	-	1.000	1.000			1.000		

189 之间,申请者是信用好的机会是 $851/63$, 其对数 s 等于 2.60 (在第 3 行的第 5 列), s' 和 s 是等价的。在采样中每个记分段人数占总数的百分数在第 10 列 (即记分的分布 $p_{s(s)}$); 在每个记分段中信用好 (或差) 的人数占信用好 (或差) 总数的百分数在第 6 (或 7) 列 (即预测假然率 $p_{s|x(s|G)}$); 而在每个记分段中, 一个申请者是信用好 (或差) 的概率在第 8 (或 9) 列 (即决策概率 $p_{x|s(G|s)}$)。

假设银行开展一年期贷款额一万元, 利率为 15% 的消费者信贷业务, 并设同期无风险的年利率是 10%。根据式 (4) 计算无差概率

$$f(\bar{s}) = \frac{(r+1)l}{g+l} = \frac{(0.15+1)l}{0.15+1} = 0.9565$$

由式 (3) 计算得信用得分临界值

$$\bar{s} = \ln \frac{f(\bar{s})}{1-f(\bar{s})} = 3.09$$

查表 1 中的第 5 列发现从第 5 行 3.140 起均大于 3.09, $s = 3.140$ 所对应的整数段 200-209, 即记分 $s' = 200$, 故 $\bar{s} = 200$ 。银行拒绝信用得分小于或等于 200 顾客的贷款申请, 接受信用得分大于 200 顾客的贷款申请。银行从信用得分不同顾客的贷款获得的期望收益和风险列在第 11 和 12 列。

5 结论

银行在利用上例数据得出各个参数进行决策时, 首先需判断决策环境与预测环境 (历史数据所处的环境) 是否相同。如果表示系统因素的各个随机变量取值发生很大的改变, 例如研究对象有很大变化, 则原来估计的概率不再适用。根据变化了的样本对象, 重新进行估计与确定。举例说明银行利用信用评分模型有利于规避信用风险, 但对此问题的研究还有待进一步深化。在贷款申请者的信用评分模型中, 假设申请者的各方面因素是相互独立的。另外, 在衡量银行的信用风险时未考虑各笔贷款间的相关性。若能确定各个贷款间的相关系数, 找出最佳的贷款组合, 则可进一步控制银行的信用风险。这对提高贷款决策的科学化水平, 规避信用风险水平都是很有意义的。

参考文献

- [1] 宋清华. 论贷款风险管理的策略. 财经问题研究. 大连, 1997
- [2] 山东城市金融学课题组国有商业银行科学贷款决策体系的研究. 当代经济科学, 1997
- [3] 试析消费者贷款及其在我国的应用. 金融科学 (中国金融学院院刊), 1997
- [4] 詹原瑞. 影响图理论和应用. 天津大学出版社, 1995
- [5] 詹原瑞. 决策与预测讲义. 天津大学管理学院, 1997
- [6] K. T. Marshall Decision Making and Forecasting, 1995 McGraw-Hill, Inc.

Model Design And Decision Analysis on Credit Rating

Zhan Yuanrui Tian Hongwei

(Management School, Tianjin University)

Abstract This paper focuses on the consumer credit problem in the credit business of commercial banks. Using decision analysis techniques, it gives the design principle of credit rating and the method about how determining the critical rate. It also provides the calculation formulas of expected return and risk in the optimum credit strategy. Finally, it explains these by an example.

Keyword Credit risk, odds, probability influence-diagram, decision analysis