

基于交易拓扑结构与风险评估的信贷策略模型

摘要

银行需要提前分析各企业的实力与信誉等级以减少信贷违约率并最大化收益,而对无抵押资产的中小微企业的信贷来说,风险评估就更为重要。本文主要通过各类数学模型对附件中的中小微企业进行突发因素下的信贷风险评估,从而为银行提供信贷策略。

第一问中我们根据附件一之中有信贷记录的 123 家企业的数据,我们采用监督学习方法,人工选取特征,运用 Logistic 多分类线性回归对企业信誉评级进行模型训练。同时利用 SVM 分类器,基于同样的特征对违约情况进行预测训练。之后,我们建立了银行期望利润最大化模型,基于企业在上下游交易网络中的重要性程度,以银行总期望利润为目标函数,算出当银行的“期望利润”最大时,得到各个公司贷款额度与利率的信贷策略。

第二问我们采用附件 1 的数据训练出的模型,代入附件 2 的 302 家无信贷记录的企业数据,得到其信誉评级和违约预测。同时构建出附件 2 企业的上下游交易网络图,计算出相应企业节点的带权度。之后,我们带入银行期望利润函数,以总额 1 亿为本金,依据带权度和风险进行利率和贷款额区间估计,同时用蒙特卡洛方法进行模拟,即可以得出在贷款额度为 1 亿元的情况下,目标函数值最大时,各个企业被分配到的贷款情况,最终给出信贷策略。

第三问我们分析突发因素对于银行放贷策略的影响。本文先将突发因素根据不同的标准进行分类,然后通过发生之后所得到的信息或者与此类事件相关的因素的变化趋势等观测值进行分析预测,最终得到一个突发事件发生的概率以及该类型的突发事件对于不同种类的公司影响。最后我们量化影响结果,给出一定的函数关系(调整函数)。从而使银行调整对于各个企业的信贷策略。

最终我们得以将附件中的企业量化出其信贷风险,并分别给到有无突发因素下银行的信贷策略,为银行向中小微企业的信贷决策提供数学模型。

关键词: Logistic 回归、SVM、银行期望利润最大化模型、交易网络拓扑图、信贷策略

一、 问题重述

银行是国民经济的重要枢纽,而信贷是银行的重要收入来源。如何把控信贷风险以减少信贷违约率,同时最大化资产收益对银行来说有着重要意义,尤其针对中小微企业的信贷决

策。由于中小微企业规模较小且缺少抵押资产，出现坏账的可能性更高，因此中小微企业的信贷风险管理就显得更加重要。

银行通常是依据信贷政策、企业的交易票据信息和上下游企业的影响力，向实力强、供求关系稳定的企业提供贷款，并可以对信誉高、信贷风险小的企业给予利率优惠。银行根据中小微企业的实力、信誉对其信贷风险做出评估，然后依据信贷风险等因素来确定是否放贷及贷款额度、利率和期限等信贷策略。某银行对确定要放贷企业的贷款额度为 10~100 万元；年利率为 4%~15%；贷款期限为 1 年，通过建立数学模型以解决下列问题：

1、对附件 1 中 123 家有信贷记录的企业的信贷风险进行量化分析，给出银行在年度信贷总额固定时对这些企业的信贷策略。

2、在问题 1 的模型基础上，对附件 2 中 302 家无信贷记录的企业的信贷风险进行量化评估，并给出银行在年度信贷总额为 1 亿元时对这些企业的信贷策略。

3、企业的生产经营和经济效益可能会受到一些突发因素影响，而且突发因素往往对不同行业、不同类别的企业会有不同的影响。综合考虑附件 2 中各企业的信贷风险和可能的突发因素（例如：新冠病毒疫情）对各企业的影响，给出该银行在年度信贷总额为 1 亿元时的信贷调整策略。

二、 问题分析

2.1、 问题一

在第一问中，根据附件一之中有信贷记录的 123 家企业的数据，银行为对各中小企业进行风险评估，需要综合考虑企业的实力和信誉等级。对此，本文联系监督学习方法，人工选取特征，运用 Logistic 多分类线性回归对企业信誉评级进行模型训练。同时我们利用 SVM 分类器，基于同样的特征对违约情况进行预测训练。结果显示都达到了不错的准确率。

之后，我们对于这些企业的经营情况进行分析，并创新了“银行期望利润最大化模型”，给出在贷款总额确定的情况下，基于企业在上下游交易网络中的重要性程度，以银行总期望利润为目标函数，考虑其与贷款额度、利率和公司流失率之间的关系，使得银行的“期望利润”最大，即在风险尽量小的情况下获得的尽可能多的年利息。

2. 2、 问题二

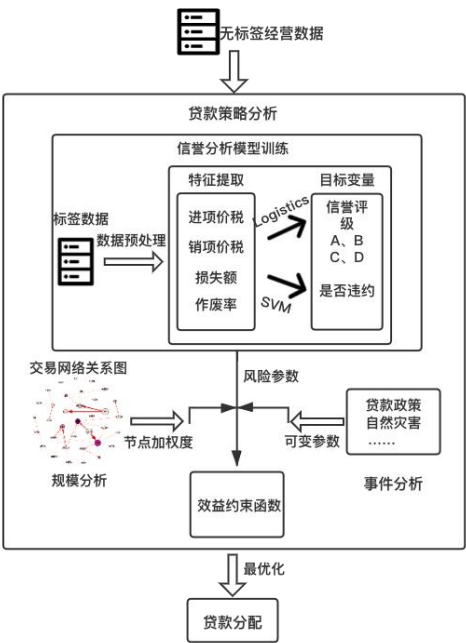
对于附件 2 中 302 家无信贷记录的企业，我们依据问题一中已训练的分类模型，对其进行信誉评级和违约预测。同时我们构建出附件 2 企业的上下游交易网络图，计算出相应企业节点的带权重。

之后，我们带入银行期望利润函数，以总额 1 亿为本金，依据带权重和风险进行利率和贷款额区间估计，同时用蒙特卡洛方法进行模拟，挑选出是银行利润最大的系数。最终完成贷款分配。

2. 3、问题三

在之前的两问之中我们建立了评分体系，考究了 A、B、C、D 四类等级的内在评价机制，以及给出了放贷策略的“目标函数”——银行期望利润函数，为了得出最优的“放贷策略”，我们使目标函数值在遵循一定的约束条件时候达到最大化。也在放贷总额为一个具体的特定值得时候，给出了对于没有信誉记录的企业信誉评级以及放贷策略。但是在之前的问题之中，我们没有考虑过任何“突发因素”，例如：新冠疫情等。我们明确突发因素的定义：突发因素是指其发生的形式无法预料，带来损失难以估计（并不是一定会带来损失，也许在某些方面其具有促进作用）的因素。

所以对于此类事件，本文通过发生之后所得到的信息或者与此类事件相关的因素的变化趋势等观测值进行分析预测，最终得到一个突发事件发生的概率以及该类型的突发事件对于不同种类的公司的影响。对于最终的影响，我们进行量化考量，给出一定的函数关系。



三、 模型构建

【第一问模型】

首先本文观察文档中的数据情况，进行数据对齐和缺失值补全，同时对可疑的交易进行了一定程度的修正。同时我们对数据进行分箱和平稳性检验，剔除异常值，从而使数据的分布更加平稳。

1.风险度量与评估预测

根据有信贷记录企业的进销数据，我们首先进行人工的特征提取，以挑选出与目标变量相关的参数，并在模型中进行验证。根据数据特点和数据量，本文根据题意计算出交易作废率 $R_1^{(E_i)}$ 、负向损失额 $N^{(E_i)}$ 、进项价税总和 $X_1^{(E_i)}$ 、销项价税总和 $x_1^{(E_i)}$ 、进销差 $M_R^{(E_i)}$ 作为我们进行分类的维度，并对数据进行标准化处理。

1.1 基于 Logistic 回归的评级分类

为对附件 1 中 123 家有信贷记录的企业进行信贷风险评估, 我们结合企业的实力与信誉等级，选择多分类 Logistic 回归的方法进行估测，将选取的指标作自变量，而将信誉评级作为因变量。具体部分相关数据如下图所示：

	0	企业代号	作废率	进项价税总和	销项价税总计	进销存	负向损失额	企业名称	信誉评级	是否违约
0	0	E1	0.055798	6.892694e+09	4.798812e+09	-2.093882e+09	-99476273.13	***电器销售有限公司	1	0
1	1	E10	0.057847	6.478611e+06	3.806986e+08	3.742199e+08	-40197.54	***建筑劳务有限公司	2	0
2	2	E100	0.001000	2.332302e+04	7.114892e+05	6.881662e+05	-3740.75	***装饰工程有限公司	4	1
3	3	E101	0.001000	5.315800e+04	7.259654e+05	6.728074e+05	1.00	***灯饰工程有限公司	4	1
4	4	E102	0.027397	3.708884e+06	4.650110e+05	-3.243873e+06	-7741.66	***大药房有限责任公司	4	1

多分类 Logistic 分析的步骤分别为：

- 第一：对模型整体情况进行说明，比如对 R 方进行描述，以及列出模型公式；
- 第二：逐一分析 X 对于 Y（相对于的对比项）影响情况；如果 X 对应的 p 值小于 0.05 则说明 X 会对 Y（相对于的对比项）产生影响关系，此时可结合 OR 值进一步分析影响幅度。
- 第三：总结分析结果。

我们将附件 1 中的数据进行预处理，分别求得不同信誉等级的企业的作废率、负向损失额、进销差、销项价税总计、 进项价税总和，代入多分类 Logistic 模型，得到下表：

多分类Logistic回归模型似然比检验						
似然比卡方值	df	p	AIC 值	BIC 值	-2LLNULL值	-2LLF值
58.305	12	0.000	308.721	350.904	337.027	278.721

这里p 值小于 0.05，因而说明拒绝原定假设，即说明本次构建模型时，放入的自变量具有有效性，本次模型构建有意义。

以信誉评级的第一项即 1.0 作为参照项进行对比分析，Y 一共有 4 项，并且以 1.0 作为参照对比项，因而最终会有 3 个公式，最终模型公式如下：

$\ln(2.0/1.0) = -357.466 + 21.344 * \text{作废率} + 377.414 * \text{负向损失额} - 38.093 * \text{进销差} + 37.652 * \text{销项价税总计} - 60.844 * \text{进项价税总和}$

$\ln(3.0/1.0) = -21.083 + 4.672 * \text{作废率} + 3.669 * \text{负向损失额} + 34.783 * \text{进销差} - 30.490 * \text{销项价税总计} - 47.870 * \text{进项价税总和}$

$\ln(4.0/1.0) = 50.282 + 22.047 * \text{作废率} + 21.022 * \text{负向损失额} - 136.183 * \text{进销差} - 415.480 * \text{销项价税总计} - 193.789 * \text{进项价税总和}$

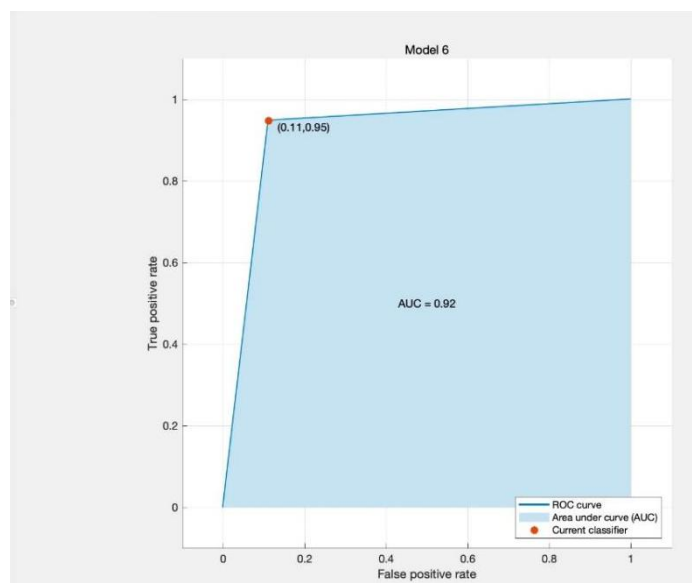
代入附件 1 的数据从而我们可以得到 123 家有信贷记录的企业的风险评估相对分数, 为下一步银行具体的信贷策略做好前提准备。

预测准确率汇总							
		预测值				预测准确率	预测错误率
		1.0	2.0	3.0	4.0		
真实值	1.0(n=27)	9	7	6	5	33.33%	66.667%
	2.0(n=38)	3	17	6	12	44.74%	55.263%
	3.0(n=34)	1	11	9	13	26.47%	73.529%
	4.0(n=24)	0	2	0	22	91.67%	8.333%
汇总						46.34%	53.66%

可以看到，模型对 D 类企业的预测准确率较高。由于 D 类企业与违约强关联，因此该模型对我们的风险判断仍有很大的指导意义。

1.2 基于 SVM 的违约分类

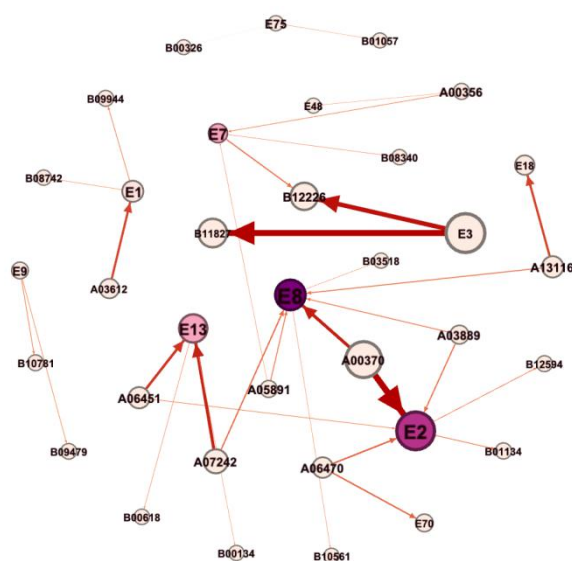
对于违约的处理，我们将上述五个变量作废率，作废率、负向损失额、进销差、销项价税总计、进项价税总和结合信誉等级，通过 SVM 进行二分类违约判断，通过 Matlab 运行出相关结果，ROC 曲线最终显示预测准确率达 97.6%。



2. 贷款基本额度分配策略

在企业经营发展过程中，关键供应链和销售链起着至关重要的作用。掌握企业上下游的关键节点，有利于我们正确认识企业在整体行业经营中所占据的生态位。这种重要性决定了企业经营的稳定性和中心地位，也应成为银行进行风险评估和资金分配的重要依据。

对于企业的上下游企业，我们可以根据所给数据构建相应的交易网络拓扑图，根据节点的带权度对节点的重要性程度进行度量，从而得到对其企业的重要性评估。相应的，带权度高的结点，说明其在网络中起中心作用，反映在拓扑图中即为更大的结点和更粗的权边，从而可以直观地展现出其企业实力。以附件 1 交易次数在 200 以上的企业上下游数据为例，我们用 Gephi 对企业的上下游交易节点进行了可视化：



贷款基本额度的分配策略如下：

若 $\frac{\text{可以贷款企业上下游权重之积}}{\text{D类企业上下游权重之积}} - 1 > 0$ ，则企业可得贷款额度：

$$\left(\frac{\text{可以贷款企业上下游权重之积}}{\text{D类企业上下游权重之积}} - 1 \right) \times (\text{贷款最高额度} - \text{贷款最低额度}) + \text{贷款最低额度}$$

若 $\frac{\text{可以贷款企业上下游权重之积}}{\text{D类企业上下游权重之积}} - 1 \leq 0$ ，则该企业有两种选择：

- (1) 不进行贷款； (2) 只贷款最低额度；

同时结合企业本身规模，对其在行业中所占比重进行加权，从而可以最终得到分配金额的结果。

3.期望利润最大化模型

接下来，为了研究究竟如何放贷才能使银行的“期望利润最大化”，我们对于已有数据进行分析，得出“基于风险评估与交易拓扑结构的期望利润最大化模型”。

首先，我们来对于“期望利润”这个概念进行解释说明：企业在经营过程中会有诸多不确定因素的存在，所以最终企业在还款时，成功还款这一行为发生的可能性的范围在 0% ~ 100%，在银行进行放贷时，无法确定该企业最终能否成功还款，所以还款的可能性即为最终的“期望”，而银行的利润即为最终通过给企业贷款所设置的利率所获得的盈利，该盈利与还款可能性相乘，即为“银行期望利润”。具体表达式如下：

$$M_i^{(\text{bank})} = W_i * L_i * \beta_i * (1 - \alpha_i)$$

此时 L_i 是变化，根据 L_i 与 α_i 的对应关系 $L_i \rightarrow \alpha_i$ ，使得 $M^{(\text{bank})}$ 最大； β_i 为还款能力函数，即对第 i 个公司银行期望利润 = 借贷给该公司的贷款 * 年利率 * 该公司还款的能力（即该企业进权与出权的乘积）*（1 - 该利率下该公司流失的可能性）

我们设 i 为每个公司， j 为交易计数器， L_i 利率先均为 0.04，

每个公司的支出 θ_i : $\sum_{j=1}^n (-X_{ij} - Y_{ij}) + \sum_{j=1}^n (-Y_{ij}) + [-W_i(1 + L_i)]$

每个公司的收入 I_i : $\sum_{j=1}^n (X_{ij} + Y_{ij}) + W_i$

根据题目要求均已一年为期，银行年度贷款总额为 W

$$W = \sum_{i=1}^n W_i \quad (n = 123)$$

W_i 为第 i 个公司贷款所得贷款。

接下来，我们做第一步的筛选，如果进销差 $(l_i - \vartheta_i) < 0$ ，即该企业无利润，则不予贷款；于是，我们仅需在进销差 $(l_i - \vartheta_i) > 0$ 的公司中进行考量，考虑其贷款额度。

最终，我们可以得知，总期望利润函数 $M^{(Bank)} = \sum_{i=1}^n M_i^{(Bank)}, (n=123)$ 即为我们需要的目标函数，为了使目标函数取得最大值，我们不妨令目标函数之中的每一个 $M_i^{(Bank)}$ 遵循相应的约束条件尽可能大，约束条件如下：

$$W_{imax} = 100, W_{imin} = 10$$

即可得到这样放贷是最好的策略，可以使期望利润达到最大。

【第二问模型】

在第一问中，我们已经基于企业的进销情况对企业的信誉评级和是否违约建立了分类模型，并结合企业的重要性评估构造出了期望利润最大化函数。针对附件 2 数据，本文将求出相应参数，并在此期望利润最大化函数约束下求出贷款额度和利率分配的最优解。

1.风险预测

根据第一问训练的分类模型，我们按照同样的数据格式，计算出相应的交易作废率、负向损失额、进项价税总和、销项价税总和、进销存。然后利用 matlab 将数据代入模型，从而输出相应的预测结果。部分预测结果数据如下图：

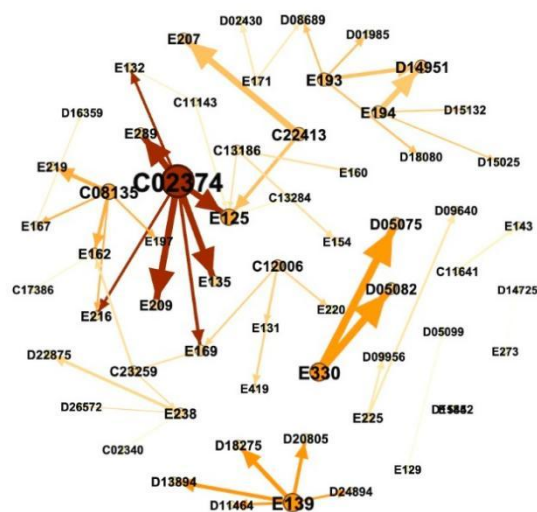
企业代号		作废率	进项价税总和	销项价税总和	进销存	负向损失额	信誉评级	是否违约
0	E124	0.143067	9.650616e+08	9.921813e+08	2.711965e+07	-2409428.80	1	0
1	E125	0.142357	1.195105e+09	1.177813e+09	-1.729178e+07	-2411467.63	3	0
2	E126	0.151771	6.503090e+08	1.340063e+08	-5.163027e+08	-517900.83	3	0
3	E127	0.027778	7.177467e+08	1.854773e+06	-7.158919e+08	-39977.94	2	0
4	E128	0.042375	2.646424e+08	1.078804e+07	-2.538543e+08	-40993.21	1	0

通过人工检视，可以发现该预测具有较高的合理性和准确率，可以当做下一步进行分析的基础。

2.构建交易网络

同样的，我们选取交易次数在 200 以上的企业，画出相应的交易拓扑图，并计算输出相应的带权度。

Id	E209	E289	E135	E125	E169	E216	E132	E419	E131	...	E143	E193	E139	E194	E225	E273	E330	E171	E129	E185
weighted degree	1535	1473	1429	3105	1545	1018	835	471	430	...	410	2253	3728	1986	904	46	3620	775	188	316



从中我们可以看到，C02374 作为最大的供应商，在节点中具有举足轻重的地位。同时 E139、E330 等企业也同样拥有广泛且稳定的客源，在市场中展现了自己的实力和地位。通过带权度，我们可以一定时间范围内量化企业的重要性，从而为银行贷款决策提供参考。

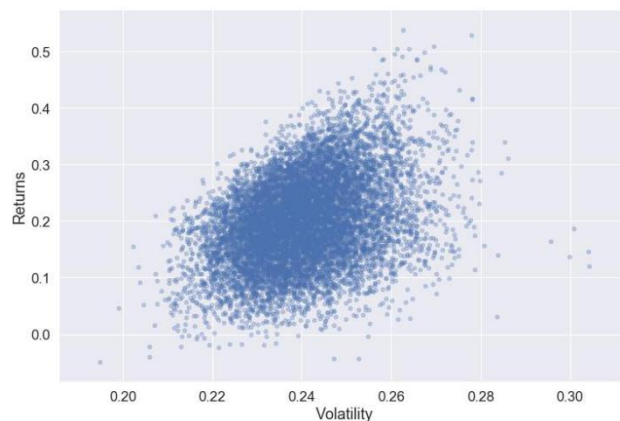
3.基于蒙特卡洛方法的函数最优解

已知“银行期望利润”。具体表达式如下：

$$M_i^{(bank)} = W_i * L_i * \beta_i * (1 - \alpha_i)$$

蒙特卡罗方法让所求解问题可以转化为某种随机分布的特征数，比如随机事件出现的概率，或者随机变量的期望值。通过随机抽样的方法，以随机事件出现的频率估计其概率，或者以抽样的数字特征估算随机变量的数字特征，并将其作为问题的解。

依据已有的企业在整体企业中重要性占比，以及得到的信誉评级和违约情况，我们对具体的利率组合进行蒙特卡洛模拟，遍历所有结果，最终可以得到最优解。对于最终得出的数据，我们进行了区间限制和调整，最终得到贷款金额分配和利率的最优组合。分配结果示例如下：



	企业代码	贷款额度	利率
0	E216	1000000.000000	0.0425
1	E132	867699.890146	0.0425
2	E419	489445.087735	0.0425
3	E131	446839.464386	0.0425
4	E220	422938.748850	0.0465

【问题三分析】

在之前的两问之中我们建立了评分体系，考究了 A、B、C、D 四类等级的内在评价机制，以及给出了放贷策略的“目标函数”——银行期望利润函数，为了得出最优的“放贷策略”，我们使目标函数值在遵循一定的约束条件时候达到最大化。也在放贷总额为一个具体的特定值得时候，给出了对于没有信誉记录的企业信誉评级以及放贷策略。但是在之前的问题之中，我们没有考虑过任何“突发因素”，例如：新冠疫情等。我们明确突发因素的定义：突发因素是指其发生的形式无法预料，带来损失难以估计（并不是一定会带来损失，也许在某些方面其具有促进作用）的因素。

所以对于此类事件, 本文通过发生之后所得到的信息或者与此类事件相关的因素的变化趋势等观测值进行分析预测, 最终得到一个突发事件发生的概率以及该类型的突发事件对于不同种类的公司的影响。对于最终的影响，我们进行量化考量，给出一定的函数关系。

1.模型建立准备

我们通过调查研究后可以得到对于突发因素分类应如何进行:

按照成因进行分类：自然性突发因素（Nature）、社会性突发因素(Social)

按照危害性进行分类：重度危害、中度危害、轻度危害

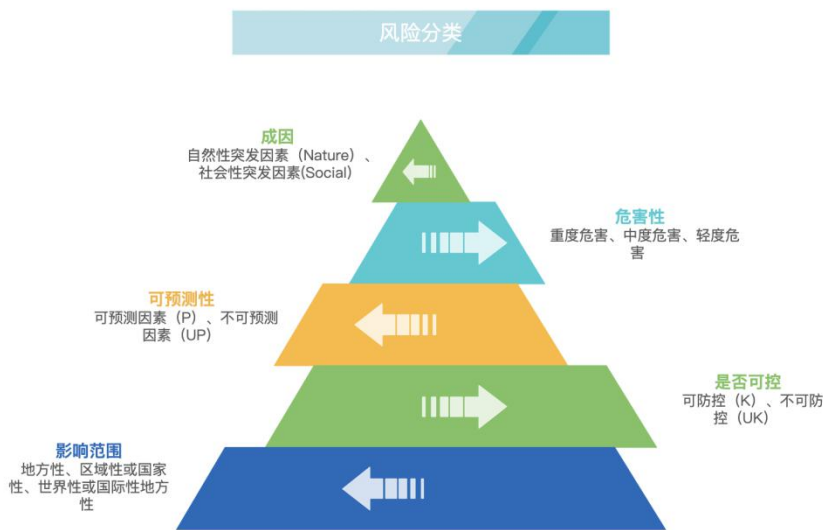
按照可预测性进行分类：可预测因素（P）、不可预测因素（UP）

按照是否可防控：可防控（K）、不可防控（UK）

按照影响范围：地方性、区域性或国家性、世界性或国际性地方性

根据我们之前所建立的函数，我们应该知道，突发因素影响的点有且只有 β_i 这一能够反映还款能力的函数，所以在突发因素发生以后，新的还款能力函数我们记为 β_i^{new} 。

在突发因素产生的时候，我们首先需要做的就是对于该事件进行定性，分类方式可以建立一决策树，如下图所示：



按照上图，最终会有 72 种情况产生，也就是说可以将突发事件分为 72 类，我们将其从左至右进行编号，不同类型的突发事件，对于不同类型的企业会产生不同程度的影响，针对于附件 2 中的企业，我们可以将其分为：农业类、基建类、医疗类与民用类，分别记为 $\{Q_i | i=1,2,3,4\}$ 。

我们依据以上所建立的决策树进行由上至下的顺序分析，再此分析过程之中我们需要考虑每一个性质对于最终 β_i^{new} 的具体影响。

2.具体影响分析

由于是突发时间，我们只能根据事件发生以后对于整体的经营情况的具体影响，根据收集到的数据不断进行回归分析，具体过程如下：

原本的 β_i 可以在这里作为一个关于时间函数的因变量，即 $\beta_i = f(t)$ （其中 t 以天作为单位，以开始有记录的时间为第 0 天），并且使用“插值”的方法使该函数变得尽量光滑；

在突发因素发生之后我们收集到的数据应当不断更新，在这些数据之中，我们不断进行回归分析与插值模拟，使得最终得到的 $\beta_i^{new} = g(t)$ 也尽量光滑；

在相同的时间段内，我们令两个函数相比，即
$$\frac{\beta_i^{new} = g(t)}{\beta_i = f(t)} = R_{\substack{\text{(企业编号)} \\ \text{(决策树上至下层数)}}} ;$$

进行至决策树的最底层，我们此时的到的 R 即为该种类突发因素对于最终还款能力的量化表达 β_i 的影响，令 $\beta_i \cdot R$ 即为新的还款能力的量化表达；

我们对于需要进行安排的企业，都计算出其 R ，接下来我们即可得到相应的调整函数：

若 $1-R < 0$ ，这时候该企业的还款能力上升，我们在原本给予其的贷款额度上进行相应比例的扩充，即 $W_i^{(new)} = R \cdot W_i$ ，扩充部分金额为 $(R-1) \cdot W_i$ ；

若 $1-R > 0$ ，这时候该企业的还款能力下降，我们在原本给予其的贷款额度上进行相应比例的缩减，即 $W_i^{(new)} = R \cdot W_i$ ，缩减部分金额为 $(1-R) \cdot W_i$ ；

3.结论分析总结

对于不同性质的突发因素，各个公司所受到的影响均有不同，最终反映到目标函数上的最直接的影响即为对于各个公司所发放的贷款额度 W_i 的变化，变化我们给出一个确定的变化系数（变化率 R ），对于每一个公司来讲，只要按照我们在模型建立与求解的部分给出的调整函数，最终将调整函数融入到银行总期望利润函数之中，即可得到新的分配方式。在这种突发因素对应的分配方式下，银行的期望利润仍是最大的，即在承担最小的风险时得到最大化的利润。