机器学习方法能识别中国系统性金融风险的概率吗？

王　达　周映雪

摘　要：本文采用梯度提升树这一机器学习模型，基于美国等17个国家的25个特征变量的宏观经济数据集，构造了风险识别模型对中国的系统性风险概率进行全面解析，并通过SHAP Value解释模型，在非线性非参数模型下探索中国的风险影响因素。实证结果表明，梯度提升树模型对系统性风险的捕捉能力显著优于传统的逻辑回归模型，其能够较好地刻画中国的风险概率走势；经过SHAP Value分解可发现，信贷因素、货币因素、金融市场化因素及国内总储蓄等是主要的风险拉动因子，且均存在明显的阈值效应。

关键词：系统性风险　机器学习　梯度提升树　SHAP Value

一、引言

近年来，全球公共部门债务规模的不断扩大和企业部门杠杆率的迅速增长，使得全球金融体系的风险水平呈上升之势。尤其是2023年以来，发展中国家不平衡的经济反弹及俄乌冲突的双重冲击下，全球金融体系面临严峻挑战。如何改善系统性风险识别的准确性和有效性，再次成为学术界和监管层关注的焦点。中国作为世界第二大经济体，其金融体系的稳定受到广泛关注。中国共产党第二十次全国代表大会的报告也强调要加强对系统性风险的监测和预警，确保不出现系统性风险。

众所周知，对金融危机的识别和预测一直是开放宏观经济学的重要研究领域。从早期预警模型如FR概率模型（Frankel & Rose, 1996）、STV横截面模型（Sachs et al., 1996）、KLR模型（Kaminsky et al., 1997），到DCSD模型（Berg & Pattillo, 1999）及逻辑回归（苏冬蔚和肖志兴，2011）等经济计量模型，学者们对于金融危机的成因和预测进行了大量的探索。2008年全球金融危机爆发后，国内外学术界在拓展微观数据、指标筛选及探索计量方法等方面，对系统性风险识别问题进行了更加深入的研究，并发展出了CoVaR条件在险值、Co-Risk模型、综合指数模型、

王达，香港中文大学（深圳）国际事务研究院客座教授，吉林大学经济学院博士生导师；周映雪，吉林大学经济学院博士研究生。本文系2022年国家社科基金重点项目“中美科技竞争的驱动逻辑、演进态势及对我国经济安全的冲击研究”（项目批准号：22AGJ006）的阶段性成果。

系统期望损失 SES 模型及银行间资产负债模型等众多方法(Bisias et al., 2012)。然而，大部分研究并未突破以探求因果关系为核心的传统经济学的研究范式，这使得对系统性风险的研究始终面临着解释性强、预测性弱的短板。

近年来，随着大数据方法的日益成熟，机器学习、深度学习等数据算法在经济学研究中得到了认可和一定程度的应用。机器学习方法不做事前的理论假设而是通过建模挖掘数据之间主要关联（Predominant Patterns）以提高预测性能。已有研究表明，机器学习模型的预测性显著优于传统方法。尤其值得注意的是，由 Friedman(2001) 提出的梯度提升树（Gradient Boosting Decision Tree，GBDT）这一主流的机器学习算法，由于对不同类型的指标包容性较好，模型泛化性能优异且在一定程度上可以解决过拟合问题，近年来受到了英格兰银行等主要国家中央银行的关注，并逐渐在宏观经济金融研究中得以应用（Bracke et al.，2019）。杨子晖 等（2022）也将梯度提升树应用到系统性风险的研究中并获得了较好的结果。Zięba et al.（2016）的研究更是指出，在统计推断方面，与传统的以线性模型为主的统计模型以及大多数机器学习模型相比，梯度提升树算法在处理诸如危机识别等二元分类问题方面的表现是最优的[①]。

为了克服机器学习模型的“黑箱机制”[②]在探究因果关系方面的不足，近年来出现了一批专门针对机器学习算法的解释模型，其中以解释向量模型（Explanation Vectors）、IME（Interactions-based Method for Explanation）模型及 LIME（Local Interpretable Model-agnostic Explanations）模型为代表。上述解释模型的出现和发展，增强了机器学习模型对于因果关系的解释能力，从而拓展了其在经济学研究中的应用性。尤其是在系统性风险分析等对预测性和解释性均具有较高要求的研究领域，机器学习算法大有可为。在众多 IME 解释模型中，由 Lundberg & Lee（2017）在合作博弈论的基础上提出的 SHAP Value 解释模型一经提出便引发了学术界的广泛关注。2017 年以来，英格兰银行开始探索机器学习方法在中央银行宏观经济金融调控中的应用；2019 年，英格兰银行在金融风险评估等领域引入了 SHAP 解释模型并取得了良好的效果（Bracke et al., 2019），这一基于量化输入影响（Quantitative Input Influence）的解释模型大大提高了非线性、

① 机器学习分为回归分析和分类分析，二者本质相同，均为对输入做出预测。区别则在于其输出空间定性或定量：若输出的是定性的离散指标，则为分类问题；若输出的是定量的连续的属性值则为回归问题。危机识别和预测的结果是典型的0-1二元分类问题，即发生危机（1）和不发生危机（0）。因此，危机识别在机器学习模型中属于典型的分类分析。

② 所谓“黑箱机制”（Blackbox Mechanism）是指，由于机器学习模型大多是非线性、非参数模型，这使得其与传统的统计模型相比，对于模型结果的解释性较差。机器学习在因果推断方面的这一短板是长期以来其难以成为社会科学主流研究方法的重要原因之一。

非参数结构的机器学习模型的可解释性。

在此背景下，本文使用梯度决策树这一主流机器学习模型并借鉴英格兰银行在研究中所使用的 SHAP Value 解释模型，以 16 个国家在 1990—2012 年[①] 期间的宏观数据和风险状况作为供梯度决策树模型学习的训练集，并以美国的同期数据作为模型的测试集，再利用训练测试完成的模型考察中国在 1990—2019 年期间的系统性风险水平及影响因素。本文更加深入的探索了“机器学习 + 解释模型”这一研究范式在中国系统性风险识别及影响因素研究中的应用，可作为其他分类和回归任务应用该研究范式的示例，同时对于探索系统性风险分析的大数据方法也具有积极的理论价值和现实意义。

二、研究方法

（一）梯度提升树模型

梯度提升树(GBDT)是Friedman(2001)提出的一种减小监督式学习（Supervised Learning）中偏差的机器学习算法，即所谓的提升（Boosting）算法。在提出之初就和支持向量机（SVM）一起被认为是泛化能力较好且适用于各种分类问题的模型。梯度提升树的本质是集成学习，其在原数据上构建多个不同的模型并汇总和集成所有模型的建模结果，因此回归或分类的表现显著优于单个模型。梯度提升树是由许多树形模型构成，其主要思路为不断建立树学习器来拟合前一个树学习器的负梯度来矫正模型的偏差，最后将每个分类器的结果相加即为最终结果。本文选择以 MSE 均方误差为损失函数的回归决策树拟合梯度，选择对数损失作为初始损失函数。

（二）模型性能评价指标

模型的风险识别性能则使用基于混淆矩阵的多维指标体系进行评价，基于混淆矩阵的多维指标体系是衡量二分类问题机器学习模型的常用指标，主要包含：第一，准确率（Accuracy），即所有预测正确的样本数目占总样本数目的比例，表示模型整体效果，越接近 1 越好；第二，召回率(Recall)，即所有真实为风险年的样本中模型预测正确的比例，表示模型捕捉风险的能力；第三，精确率（Precision），即所有预测为风险年的样本中真实风险年所占比例；第四，F1 分值（F1-Score），即召回率和精确率的调和平均数，越接近 1 越好，是衡量模型整体表现的重要指标。

此外，还有 ROC 曲线及其下方 AUC 面积用以衡量模型整体表现。ROC 曲线表示在尽量捕捉少数类时多数类如何变化，以混淆矩阵中的假正率[②] 为横坐标，召回率为纵坐标，ROC 曲线通常为凸，越靠近左上方越好，表示捕捉更多的少数类但是多数类判错很少。ROC 曲线下方的面积称

① 其中匈牙利由于受2008年肇始于美国的全球金融危机冲击较为严重，为尽可能弱化美国因素对训练集数据的影响从而更好地测试模型，匈牙利特征数据的时间跨度被设定为1990—2006年。

② 假正率（FPR）表示在所有预测为安全年的样本中真实风险样本所占的比例。

为 AUC，AUC 越大，ROC 曲线越靠近左上方，模型对少数类的分类表现也就越好。

（三）SHAP Value 解释模型

SHAP Value 解释模型是由 Lundberg & Lee（2017）开发的机器学习解释模型，其基础是 Shapley（1953）提出的合作博弈论。与传统机器学习模型中的 feature_importance 接口相比，SHAP Value 可计算单个样本中各特征的贡献值重要性，且还可以直接确定其影响方向。假设第 j 个样本点第 i 个特征的 SHAP Value 为：ψ_i^j，样本的各特征 SHAP Value 总和与基准值相加则可得到对应样本预测值，计算公式如下：

$$y_j = E(\hat{y})+\psi_1^j+\psi_2^j+\cdots+\psi_N^j \quad (1)$$

当 $\psi_i^j>0$ 则说明 i 特征对于 j 样本来说有正向作用，反之则为反向作用。而每个特征的 SHAP Value 本质是各个特征对模型结果累加边际贡献的均值。计算公式为：

$$\psi_i^j=\frac{1}{N!}\sum_{S\subseteq N\{i\}}|S|!\,(|N|-|S|-1)!\,[v(S\cup\{i\})-v(S)] \quad (2)$$

其中 N 为整体特征数，$v(\cdot)$ 表示所有特征施加影响后的最终值，如 $v(N)$ 表示 N 个特征共同作用后的终值。

更为重要的是，每个特征的 SHAP Value 即其对预测值的影响程度可以通过可视化的作用力图进行表示，从而极大地增强了该解释模型的直观性和可读性。

三、数据和结构说明

参考范小云（2006）和 Laeven & Valencia（2013），选择在 1990—2002 年期间发生严重系统性风险（风险损失超过 GDP 的 10%）或多次发生风险的国家[①] 作为训练样本，共计筛选出 16 个国家，包括韩国、马来西亚、泰国、土耳其、墨西哥、巴拉圭、委内瑞拉、芬兰、日本、厄瓜多尔、牙买加、保加利亚、匈牙利、巴西和阿根廷及印度尼西亚。本文通过对训练样本的交叉验证确定模型的超参数，另以美国作为测试样本，对模型的泛化性能进行测试，训练和

表 1　风险样本

国家	风险年份
韩国	1997—2002
马来西亚	1997—2002
泰国	1997—2002
土耳其	2000—2002
巴拉圭	1995—1999
委内瑞拉	1994—1995
芬兰	1991—1995
日本	1991—2002
厄瓜多尔	1995—2001
牙买加	1994—2000
保加利亚	1994—1996
匈牙利	1991—1995
印度尼西亚	1997—2002
巴西	1994，1995—1999
阿根廷	1995，2001—2002
墨西哥	1991，1994—1997
美国	2007—2013

① 风险定义及风险损失的计算详情可见范小云（2006）和Laeven & Valencia（2013）。

测试样本的风险年份总结见表 1。最后再利用训练完成的模型对中国的系统性风险进行全面的解析。需要注意的是，为更好地确定最终模型能否在训练时间之外准确识别风险，上述训练集和测试集的样本时间为 1990—2012 年，而中国的样本时间为 1990—2019 年。

对于风险标识，需要进行标签的二值化处理，即风险年份标识为 1，否则标识为 0。并且，由于风险标签具有时间属性，我们希望风险在发生之前即被识别，所以我们选择 lead-lag 方法对风险标签进行前置，如 lead-lag1 表示将第 t-1 年标签为风险年，以表示第 t 年发生系统性风险这一事件（即风险识别标识前置 1 期）。具体前置几期会对风险有更好的识别效果，后续会通过实验来确定。

对于供机器学习模型训练的特征，通过总结相关文献，在数据可得性的基础上确定了 25 个指标，涵盖了宏观经济金融的

表 2　特征说明

特征类别	特征名称	特征符号	数据来源
国内银行	银行国内授信 /GDP	DC/GDP	Wing 数据库
	净国内信贷现价美元	NetDC	世界银行 WDI 数据库
	对私人部门的信贷 /GDP	PC/GDP	Wing 数据库
	国内总储蓄 /GDP	TDS/GDP	世界银行 WDI 数据库
国内证券	股票交易总额	TST	世界银行 WDI 数据库
	上市公司总市值 /GDP	TMV/GDP	Wing 数据库
国际储备	非黄金总储备	TR	国际货币基金组织 IFS 数据库
国际贸易	外国直接投资净流入	FDI	世界银行 WDI 数据库
	货物和服务进口	IPT	世界银行 WDI 数据库
	货物和服务出口	EPT	世界银行 WDI 数据库
	商业服务出口额	CSE	世界银行 WDI 数据库
国际收支	经常账户余额 /GDP	CA/GDP	国际货币基金组织 WEO 数据库
	净再分配收入	NetRI	世界银行 WDI 数据库
经济增长	工业增加值年增长率	IG%	世界银行 WDI 数据库
	人均 GDP 年增长率	GDPGPC%	世界银行 WDI 数据库
	资本形成总额 /GDP	TCF/GDP	Wing 数据库
	GDP 年增长率	GDPG%	世界银行 WDI 数据库
	年通胀率	IFLA%	世界银行 WDI 数据库
国家货币	M2 同比增幅	M2%	Wing 数据库
	M2/ 准备金	M2/Reserve	Wing 数据库
	M2/GDP	M2/GDP	Wing 数据库
居民生活	政府支出 /GDP	GFCE/GDP	世界银行 WDI 数据库
	CPI 同比增长率	CPI%	Wing 数据库
	失业率	UEM%	国际货币基金组织 WEO 数据库
	最终消费支出 /GDP	FCE/GDP	世界银行 WDI 数据库

表 3　不同 lead-lag 期限下 Logit 模型与梯度提升树模型（GBDT）表现对比

lead–lag	TrainAccuracy		Accuracy		整体 F1–score		风险 Recall		Auc 面积	
	GBDT	Logit	GBDT	Logit	GBDT	Logit	GBDT	Logit	GBDT	Logit
1	1	0.7983	0.8261	0.5652	0.83	0.56	0.7143	1	0.9643	0.9196
2	1	0.8127	0.5217	0.3478	0.52	0.18	0.7500	1	0.8083	1
3	1	0.8271	0.5652	0.3913	0.58	0.37	0.6250	0	0.7333	0.0833
4	1	0.8213	0.6522	0.6522	0.66	0.51	0.7500	0	0.8083	0.1917

多个维度，总结为表 2。由于机器学习遵循数据导向逻辑，因此无须预设各特征对系统性风险发生概率的作用方向，这最后会通过解释模型来确定。

四、模型建立与评价

首先，可以对比传统逻辑回归模型与梯度提升树模型的表现，并探索最优的风险前置期；其次，调整模型超参数优化模型的泛化能力，最后应用训练好的模型进行中国风险识别，所用模型的代码均基于 Python3.7。

（一）模型对比和风险前置期确定

首先，固定模型初始参数对比 Logit 模型和梯度提升树的表现。Logit 模型需要对数据进行标准化，选择 Z-score 形式，而梯度提升树模型对数据分布不甚敏感可直接使用原始数据建模；为保证模型的充分学习，Logit 模型中的迭代次数设为 10000，梯度提升树中树的数目设为 500。基于此，分别在标签 lead-lag1~4 年下建立两种模型并将模型表现总结为表 3。

从训练集表现对比，梯度提升树模型经过训练后，对风险的识别度都达到了 100%[①]，显著高于 Logit 模型；从测试集数据来看，lead-lag 1 期的梯度提升树模型对风险的识别精度为 82.6%，不仅显著高于 lead-lag 1 期 Logit 模型的 56.5%，而且在所有模型中的表现是最好的。从训练集的召回率、准确率及 F1 分值这三项指标来看，lead-lag 1 年的梯度提升树模型的表现也基本上是最优的。

因此，可以得到两个基本结论。首先，在风险预测精度方面，梯度决策树模型显著优于 Logit 模型；其次，lead-lag1 年的梯度提升树模型优于超前 lead-lag2~4 年的三个模型。所以，本文选择 lead-lag1 年的梯度提升树模型进行后续调整。

（二）模型超参数调整

在模型超参数优化中，选择交叉验证和网格搜索这两种机器学习模型常用的方法。首先对学习率（learning_rate）和树的数目（n_estimators）进行综合网格搜索，

① 由于机器学习算法复杂度较高，因此在不调校模型参数的情况下往往会出现过拟合现象。

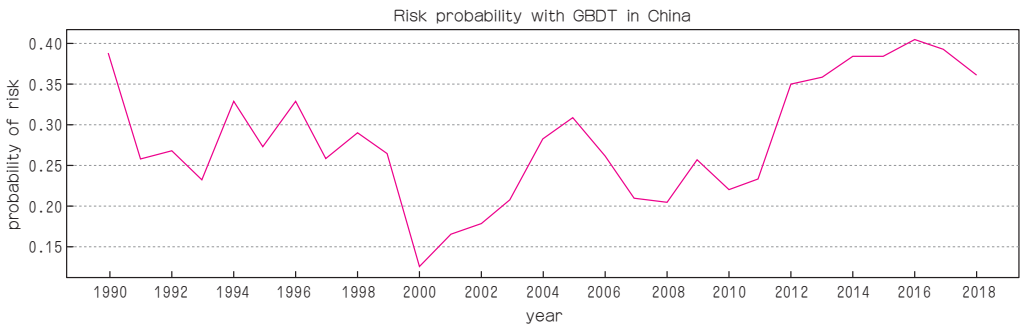

图 1　1991—2019 年间中国系统性风险概率图

根据训练集交叉验证的结果，确定了在学习率为 0.05 且构建 90 棵树时模型效果最佳。在此基础上，对四个剪枝参数共同构建第二次网格搜索，最终确定内部节点再划分所需最小样本数（min_samples_split）为 16，叶结点最小样本数（min_samples_leaf）为 3，树的最大深度（max_depth）为 5 以及划分结点时考虑的最大特征数（max_features）为 4，最后在现有参数组合下再调整学习率到 0.02。此时，训练集上的五折交叉验证得分均值较高，为 0.787。

建立上述确定的模型并进行测试可得，测试集整体的预测准确率为 0.826，训练集上的准确率为 0.936，过拟合现象已大大改善。计算 AUC 面积为 0.804，表示模型有一定的分类能力和泛化能力。此外，模型对风险年的召回率为 0.714，表示在 1991—2013 年美国所发生的风险中模型可准确识别其中的 71.4%。通过模型的混淆矩阵可知，模型对于 1991—2013 年美国系统性风险的预测中，未发生风险预测准确的有 14 年，发生风险预测准确的有 5 年，未发生风险但预测为风险的有 2 年，发生风险但未被预测到的也有 2 年，整体错分率是 0.1739。整体来看，加权整体准确率、召回率和 F1-score 均为 0.85，模型结果较平衡且分类结果较好。

（三）中国系统性风险概率识别

最后，利用确定好的模型识别中国的系统性风险概率。根据识别结果，做出中国 1991—2019 年间的年度风险概率图（图 1）。

从整体上看，中国的系统性风险概率一直低于 45%，说明中国并没有遭受真正的系统性风险。但是，在 1993—1994 年、1996—1998 年、2002—2005 年、2008—2009 年和 2012—2016 年等都出现了风险的阶段性高点。

结合中国的现实情况分析可知，中国在 1994 年遭遇了改革开放之后首次较为严

重的通货膨胀，但并未对经济造成严重冲击；而 1997 年东亚金融危机爆发，对中国的经济和金融都造成一定冲击，但在中国积极的财政和货币政策干预下，风险迅速得到缓解；2001 年之后，中国银行业不良贷款率较高的问题比较突出，并且“非典”的爆发也使实体经济受到一定的冲击；2008 年，全球金融危机爆发，但鉴于中国“四万亿”刺激计划的实施，这次风险对中国产生了较为有限的影响（李敏波和梁爽，2021）；到 2011 年后，中国金融体系出现流动性紧张、影子银行快速发展等问题，隐含的系统性风险概率不断上升；2015 年和 2016 年，由于中长期贷款供应增加以及房价的快速上涨，中国信贷额急速扩张，使得在 2016 年，中国的风险概率达到了 1990 年以来的历史高点。综合模型识别的结果来看，经过训练的模型可以较为准确地模拟中国系统性风险年度变化的总体趋势，证明了模型在中国数据上的有效性，也说明模型具有良好的泛化能力。

五、模型解释

在上述模型的基础上，应用 SHAP Value 计算各个特征对中国系统性风险概率的影响，之后可以从时间维度和特征维度两个方面切入，研究风险的影响因子。

首先计算中国在 1991—2019 年爆发系统性风险概率的年度平均值，作为基准值（Base Value），再计算每个特征在对应年份中对中国系统性风险概率的贡献度，即 SHAP Value。

（一）时间维度：特征贡献度年度变化分析

整理各个特征贡献度并做年度折线图如图 2 所示。由图 2 可得以下结论：第一，1994 年对中国系统性风险概率影响最大的两个特征为通货膨胀和上市公司总市值，并且通货膨胀对风险的贡献在 1994 年达到峰值（为 0.0315），之后的十多年里影响均不明显，而上市公司总市值在 21 世纪之前，其对风险概率推动力即都较为可观，到 2005 年后对风险概率的贡献则显著转为负，这说明金融改革对中国的风险管理具有重要意义，也间接说明了改革具有一定成效。第二，国内总储蓄一直以来对风险都有一定的正向拉动作用，即显著大于 0，且拉动作用不断增强，如 2000 年，国内总储蓄对风险的贡献度不到 1%（为 0.0051），到 2016 年已上升超过 5%（为 0.0527），说明当时中国流动性过剩的问题日益严重，亟待解决。第三，从 2012 年开始信贷指标对风险的贡献度迅速提高，成为推高中国系统性风险概率的主要因素，特别是国内净信贷，其对风险的贡献度在 2012 年之后一直维持在 10% 以上，私营部门的信贷 /GDP 也有相当大的贡献。第四，在整个样本期间，外国直接投资和国际贸易指标都是降低中国风险概率的重要特征，如外国直接投资和进出口的 SHAP Value 持续为负。其他特征对中国风险的影响变动较小，不再赘述。

（二）特征维度：特征贡献度与特征取值的关系分析

进一步的，可以根据 SHAP Value 及其

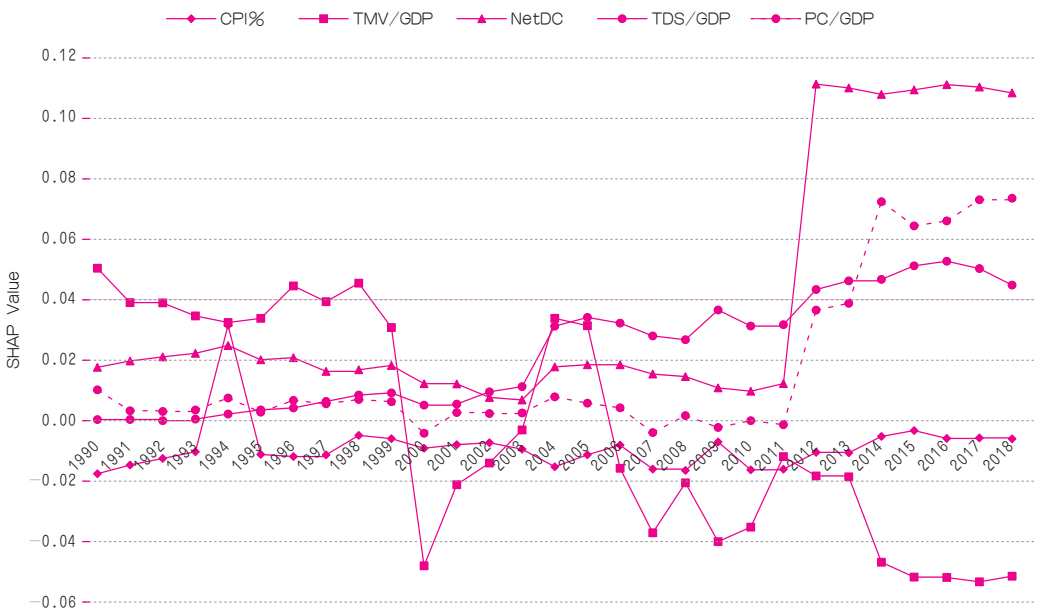

图 2　部分指标历年 SHAP Value 变化折线图

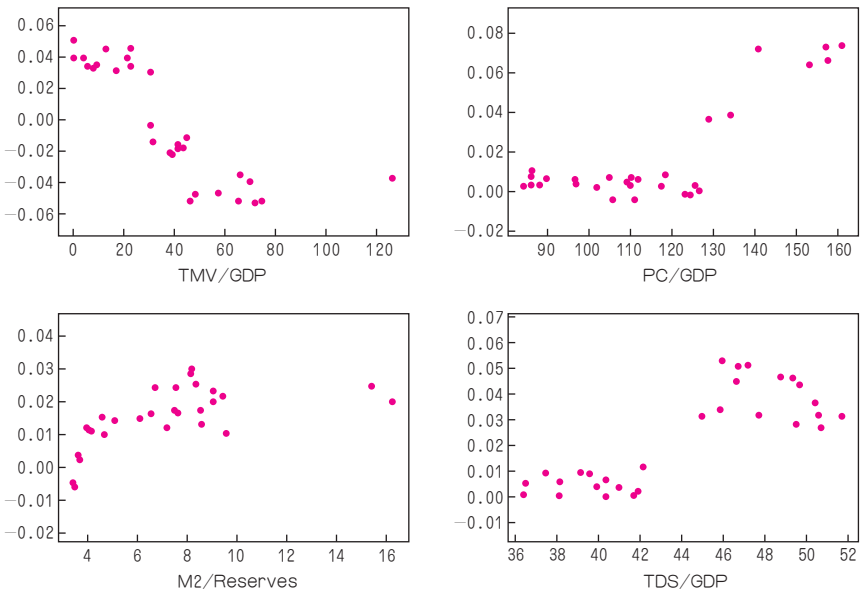

图 3　SHAP Value 散点图

对应特征值做散点图来探索二者的关系。

图 3 依次为上市公司总市值 /GDP、对私人部门的信贷 /GDP、M2/ 准备金及国内总储蓄 /GDP 的 SHAP Value 散点图[①]。由图 3 可知，对中国系统性风险概率有重大影响的几个指标存在明显的阈值效应。国内净信贷、对私人部门的信贷 /GDP 和 M2/ 准备金的阈值分别为 10 万亿人民币、130% 和 4%，一旦超过自身阈值，相应指标对风险概率的拉动作用会急剧提升。国内总储蓄 /GDP 也类似，其阈值为 42%。值得注意的是，上市公司总市值 /GDP 的门槛值为 30%，但超过这个门槛，其对风险概率的推动力会大大降低。

六、结论

本文将梯度提升树这一主流机器学习模型运用到系统性风险识别这一问题中，通过对阿根廷等 16 个国家在 1990—2012 年期间的 25 个特征组成的样本数据进行训练，并辅助以美国在 1990—2012 年期间的相同特征数据进行测试，构建了一个系统性风险识别模型。根据此模型，对中国在 1991—2019 年的系统性风险概率情况进行了全面分析。结果表明：机器学习模型在预测精度、泛化性能以及对系统性风险的捕捉能力方面均优于逻辑回归模型，且所建立的模型可较为准确地捕捉到中国系统性风险概率的变化趋势。

本文还引入近年来备受关注的机器学习解释模型 SHAP Value，对解释性较弱的梯度提升树模型进行分解。该解释模型可根据识别模型准确计算所有特征变量在不同年份对年度系统性风险概率的贡献度，并能够从时间维度和特征维度两个角度直观地展示全部特征变量对系统性风险概率的影响。实证结果表明，信贷因素、货币因素、金融市场化因素以及国内总储蓄等对中国的风险概率存在较强的拉动力，而国际贸易、外国直接投资等因素可显著降低中国的系统性风险概率；此外，对中国风险概率具有较大拉动力的几个指标在特征维度上存在较为明显的阈值效应，即超过阈值对风险概率的贡献会发生改变。整体而言，梯度提升树模型与 SHAP Value 的结合，既可较为准确的识别风险，又能精确地量化各特征对风险的影响，可大大提高实施宏观审慎政策的前瞻性和针对性，具有较为突出的政策含义。N

学术编辑：卢超群

① 篇幅原因，其他指标的散点图略。

参考文献
[1] 范小云.繁荣的背后——金融系统性风险的本质、测度与管理[M].北京:中国金融出版社,2006.
[2] 苏冬蔚,肖志兴.基于亚洲六国宏观数据的我国金融危机预警系统研究[J].国际金融研究,2011(6):14-24.
[3] 杨子晖,张平淼,林师涵.系统性风险与企业财务危机预警——基于前沿机器学习的新视角[J].金融研究,2022(08):152-170.
[4] 李敏波,梁爽.监测系统性金融风险——中国金融市场压力指数构建和状态识别[J].金融研究,2021(06):21-38.
[5] Bisias D, Flood M, Lo A W, et al. A survey of systemic risk analytics[J]. Annu. Rev. Financ. Econ., 2012, 4(1): 255-296.
[6] Berg A, Pattillo C. Predicting currency crises:: The indicators approach and an alternative[J]. Journal of international Money and Finance, 1999, 18(4): 561-586.
[7] Bracke P, Datta A, Jung C, et al. Machine learning explainability in finance: an application to default risk analysis[J]. 2019.
[8] Frankel J A, Rose A K. Currency crashes in emerging markets: An empirical treatment[J]. Journal of international Economics, 1996, 41(3-4): 351-366.
[9] Friedman J H. Greedy function approximation: a gradient boosting machine[J]. Annals of statistics, 2001: 1189- 1232.
[10] Kaminsky G, Lizondo S, Reinhart C M. Leading indicators of currency crises[J]. Staff Papers, 1998, 45(1): 1-48.
[11] Laeven L, Valencia F. Systemic banking crises database[J]. IMF Economic Review, 2013, 61(2): 225-270.
[12] Lundberg S M, Lee S I. A unified approach to interpreting model predictions[J]. Advances in neural information processing systems, 2017, 30.
[13] Sachs J, Tornell A, Velasco A. The Mexican peso crisis: Sudden death or death foretold?[J]. Journal of international economics, 1996, 41(3-4): 265-283.
[14] Shapley L S. A value for n-person games[J]. Classics in game theory, 1997, 69.
[15] Zi ba M, Tomczak S K, Tomczak J M. Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction[J]. Expert systems with applications, 2016, 58: 93-101.

Can Machine Learning Identify the Probability of Systemic Risk in China?

WANG Da[1,2]　ZHOU Yingxue[2]

(1.The Institute of International Affairs, The Chinese University of Hong Kong, Shenzhen;
2.School of Economics, Jilin University)

Abstracts　This paper constructs a risk identification model to analyze the probability of systemic risk in China using the gradient boosting decision tree with macroeconomic data comprised of 25 features from 17 countries, including the United States. It also examines the risk-drivers in China on the basis of a non-linear non-parametric model using SHAP values. The results show that the gradient boosting decision tree outperforms traditional logistic regression in capturing risk, and could simulate the general trend of the probability of systemic risk in China. Moreover, following the decomposition of SHAP value, it can be found that credit, monetary factors, financial marketization and gross domestic savings are important risk-drivers, and all have a significant threshold effect.

Keywords　Systemic Risk, Machine Learning, Gradient Boosting Decision Tree, SHAP Value

JEL Classification　C61　E44　F37