Computer Science and Application 计算机科学与应用, 2025, 15(11), 19-32
Published Online November 2025 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2025.1511279

Hans汉斯

基于WOE-Logistic信用评分卡模型构建与应用研究

孙　娜，刘政永*

河北金融学院河北省金融科技应用重点实验室，河北　保定

收稿日期：2025年9月29日；录用日期：2025年10月26日；发布日期：2025年11月3日

摘　要

本研究基于Give Me Some Credit数据集，开发了一种融合WOE编码与Logistic回归的信用评分卡模型，旨在解决金融机构在信贷风险评估中的核心挑战。研究的主要贡献在于：提出了一种优化的特征离散化方法，通过WOE转换有效处理非线性关系并增强模型解释性；构建了包含KS统计量、PSI稳定性和多决策阈值的综合评估体系，显著提升了模型验证的全面性与业务适用性。实证结果表明，该模型在测试集上取得了0.85的AUC值和0.452的KS统计量，展现出优秀的风险区分能力，同时PSI指标验证了模型在不同群体间的稳定性。本研究的方法论框架不仅为信用风险评估提供了技术参考，其评估体系也可推广至其他金融风险预测场景。然而，研究在特征工程深度和模型对比广度方面仍存在改进空间，为后续研究指明了方向。

关键词

信用评分卡，WOE，Logistic回归

Research on the Construction and Application of WOE-Logistic Credit Scorecard Model

Na Sun, Zhengyong Liu*

Hebei Key Laboratory of Financial Technology Application, Hebei Finance University, Baoding Hebei

Received: September 29, 2025; accepted: October 26, 2025; published: November 3, 2025

Abstract

This study is based on the Give Me Some Credit dataset and has developed a credit scoring card

*通讯作者。

文章引用: 孙娜, 刘政永. 基于 WOE-Logistic 信用评分卡模型构建与应用研究[J]. 计算机科学与应用, 2025, 15(11): 19-32. DOI: 10.12677/csa.2025.1511279

model that integrates WOE encoding with logistic regression, to address the core challenges faced by financial institutions in credit risk assessment. The main contributions of this research are as follows: it proposes an optimized feature discretization method which can effectively deal with nonlinear relationships and enhance model interpretability by using WOE transformation; it constructs a comprehensive evaluation system that includes KS statistics, PSI stability, and multiple decision thresholds, significantly improving the comprehensiveness and business of model validation. Empirical results show that the model achieves an AUC value of 0.85 and a KS statistic of 0.452 on the test, demonstrating excellent risk differentiation capabilities, while the PSI indicator verifies the stability of the model across different groups. The methodological framework of this study not only provides technical references credit risk assessment but also its evaluation system can be generalized to other financial risk prediction scenarios. However, there is room for improvement in the depth of feature engineering and the breadth of model, pointing the way for further research.

Keywords

Credit Scorecard, WOE, Logistic Regression

1. 引言

信用评分技术是一种对贷款申请人(信用卡申请人)做风险评估分值的统计模型。信用评分卡模型是一种成熟的预测方法，其在信用风险评估以及金融风险控制领域得到广泛应用[1] [2]。信用评分卡可以根据客户提供的资料、客户的历史数据、第三方平台数据(芝麻分、京东、微信等)，对客户的信用进行评估[3]。信用评分卡的建立是以对大量数据的统计分析结果为基础，具有较高的准确性和可靠性。

本文是利用 Python 语言，通过可视化的 Jupyter Notebook (Anaconda3)开发环境，通过对 Kaggle 上的 Give Me Some Credit 数据的挖掘分析，结合信用评分卡的建立原理，分别完成数据处理、特征变量选择、变量 WOE 编码离散化、Logistic 回归模型开发评估、信用评分卡和自动评分系统创建等步骤，构建 WOE-Logistic 信用评分卡模型来对贷款违约风险进行预测分析，为银行等金融机构进行客户信贷风险控制给予参考和指导。

2. 数据说明

本文数据是从 kaggle 上 Give Me Some Credit 获取的，可在官网(https://www.kaggle.com/)上下载数据，有 15 万条的训练数据和 10 万条未标记数据[4]。基于借贷的场景，确定“违约”的定义。根据新的 Basel Capital Accord (巴塞尔资本协议)，一般逾期 90 天算做违约，数据变量解释具体见表 1。

数据属于个人消费类贷款，只考虑信用评分最终实施时能够使用到的数据，主要从如下一些方面获取数据。

- 基本属性：包括了借款人当时的年龄。
- 偿债能力：包括了借款人的月收入、负债比率。
- 信用往来：两年内 30~59 天逾期次数、两年内 60~89 天逾期次数、两年内 90 天或高于 90 天逾期的次数。
- 财产状况：包括了开放式信贷和贷款数量、不动产贷款或额度数量。

- 贷款属性：暂无。
- 其他因素：包括了借款人的家属数量(不包括本人在内)。
- 时间窗口：自变量的观察窗口为过去两年，因变量表现窗口为未来两年。

Table 1. Table of original variables
表 1. 原始变量表

变量名	描述
Serious Dlqin2yrs	是否有超过 90 天或更糟的逾期拖欠
Revolving Utilization Of Unsecured Lines	贷款以及信用卡可用额度与总额度比例
age	借款人当时的年龄
Number Of Time 30~59 Days Past Due Not Worse	35~59 天逾期但不糟糕次数
Debt Ratio	负债比率
Monthly Income	月收入
Number Of Open Credit Lines And Loans	未偿还贷款数量和信贷额度
Number Of Times 90 Days Late	借款人逾期 90 天或以上的次数。
Number Real Estate Loans Or Lines	不动产贷款或额度数量
Number Of Time 60~89 Days Past due Not Worse	借款人已超过 60~89 天的次数，但在过去两年中没有更糟。
Number Of Dependents	家庭中的家属人数(配偶，子女等)

3. 建模分析

3.1. 数据获取

3.1.1. 加载相应程序包

本文的需要加载的程序包有科学计算、作图和机器学习等三个方面。在实践过程建议根据需要随时加载。如果加载失败，请使用 aconda 中 pipinstall 程序包名进行安装再加载。

3.1.2. 导入数据并查看基本信息

数据来源于 Kaggle 的 Give Me Some Credit 竞赛项目，其中 cs-training.csv 文件有 15 万条的样本数据，包含了 11 个变量，大致情况如表 2 所示。

Table 2. Give Me Some Credit data variable information overview table
表 2. Give Me Some Credit 数据变量信息一览表

Variable Name	Description	Type
Serious Dlqin2yrs	Person experienced 90 days past due delinquency or worse	Y/N
Revolving Utilization Of Unsecured Lines	Total balance on credit cards and personal lines of credit except real estate and no installment debt like car loans divided by the sum of credit limits	percentage
age	Age of borrower in years	integer

续表

Number Of Time 30~59 Days Past Due Not Worse	Number of times borrower has been 30~59 days past due but no worse in the last 2 years.	integer
Debt Ratio	Monthly debt payments, alimony, living costs divided by monthy gross income	percentage
Monthly Income	Monthly income	real
Number Of Open Credit Lines And Loans	Number of Open loans (installment like car loan or mortgage) and Lines of credit (e.g. credit cards)	integer
Number Of Times 90 Days Late	Number of times borrower has been 90 days or more past due.	integer
Number Real Estate Loans Or Lines	Number of mortgage and real estate loans including home equity lines of credit	integer
Number Of Time 60~89 Days Past Due Not Worse	Number of times borrower has been 60~89 days past due but no worse in the last 2 years.	integer
Number Of Dependents	Number of dependents in family excluding themselves (spouse, children etc.)	integer

3.2. 数据预处理

3.2.1. 缺失值处理

从上面数据信息中可以看出缺失值情况，我们也可以利用 isnull 函数查看一下。特征量 Monthly Income 缺失数量为 29,731 个，缺失值较多；特征量 Number Of Dependts 缺失为 3924 个，缺失值较少。

由于 Monthly Income 存在缺失值，缺失值又不是极多，因此选择填充的方法进行处理，在这里利用随机森林法，将有缺失值的变量分成已知特征和未知特征(仅含有缺失值)，将已知特征和标签进行训练，得到训练模型，对未知特征进行预测。

由于 Dependents 变量缺失值比较少，所以直接删除对总体模型也不会造成太大的影响。对缺失值处理完成后，删除重复项。

经过填补和删除缺失值后，所有特征的数据量都为 145,563。

3.2.2. 异常值处理

异常值是指明显偏离大多数抽样数据的数值，比如个人客户的年龄小于 0 时，通常认为该值为异常值。在本数据集中，采用箱线图进行分析。

(1) 信用往来

由图 1 知，信用往来的三个变量都有离群值，查看各个特征离群值的数量，它们离群值的数量都比较少，则把这些离群值都删除。其他变量离群值不是很明显，暂时不做处理。

(2) 年龄(age)

由图 2 知，age 为 0 的样本，明显是不符合常识的，应同样作为异常值舍弃。

(3) 异常值处理

剔除 age 为 0 的样本；以及剔除信用往来三个变量的的异常值，剔除其中一个变量的 96、98 值，其他变量的 96、98 两个值也会相应被剔除。

数据集中好客户为 0，违约客户为 1，考虑到正常的理解，能正常履约并支付利息的客户为 1，所以我们将其取反。

(4) 数据切分

为了验证模型的拟合效果，需要对数据集进行切分，分成训练集和验证集。

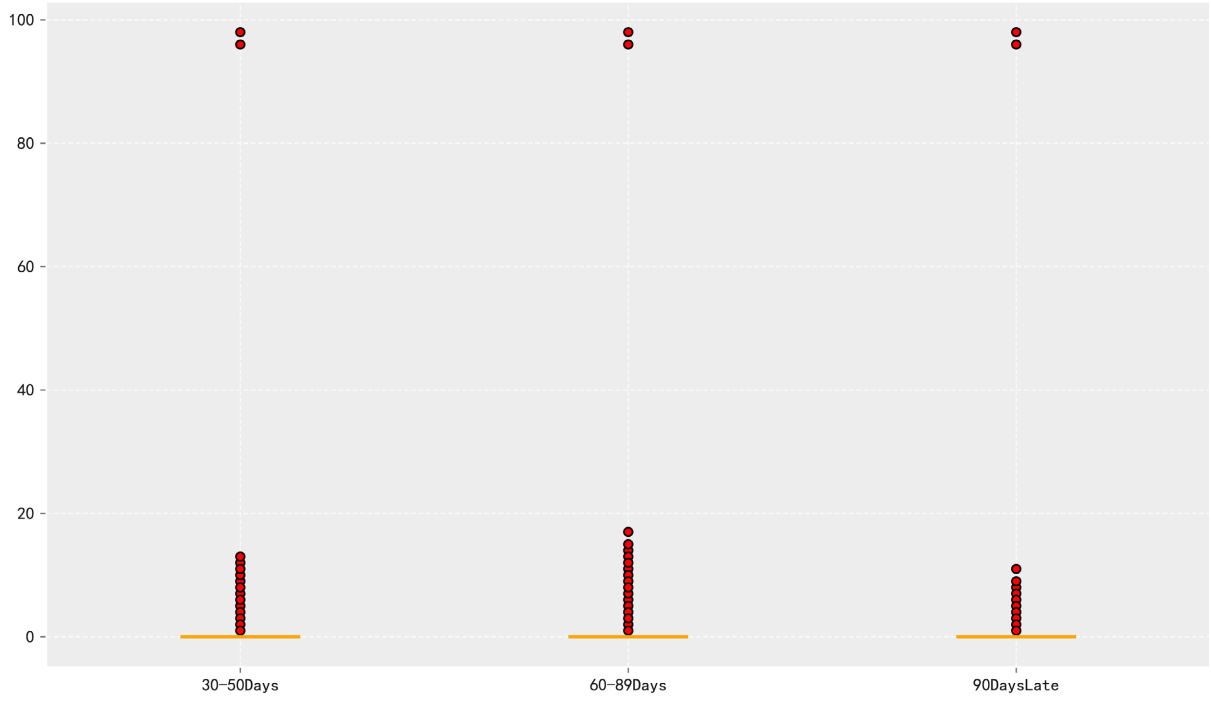

Figure 1. Credit transactions box plot
图 1. 信用往来箱线图

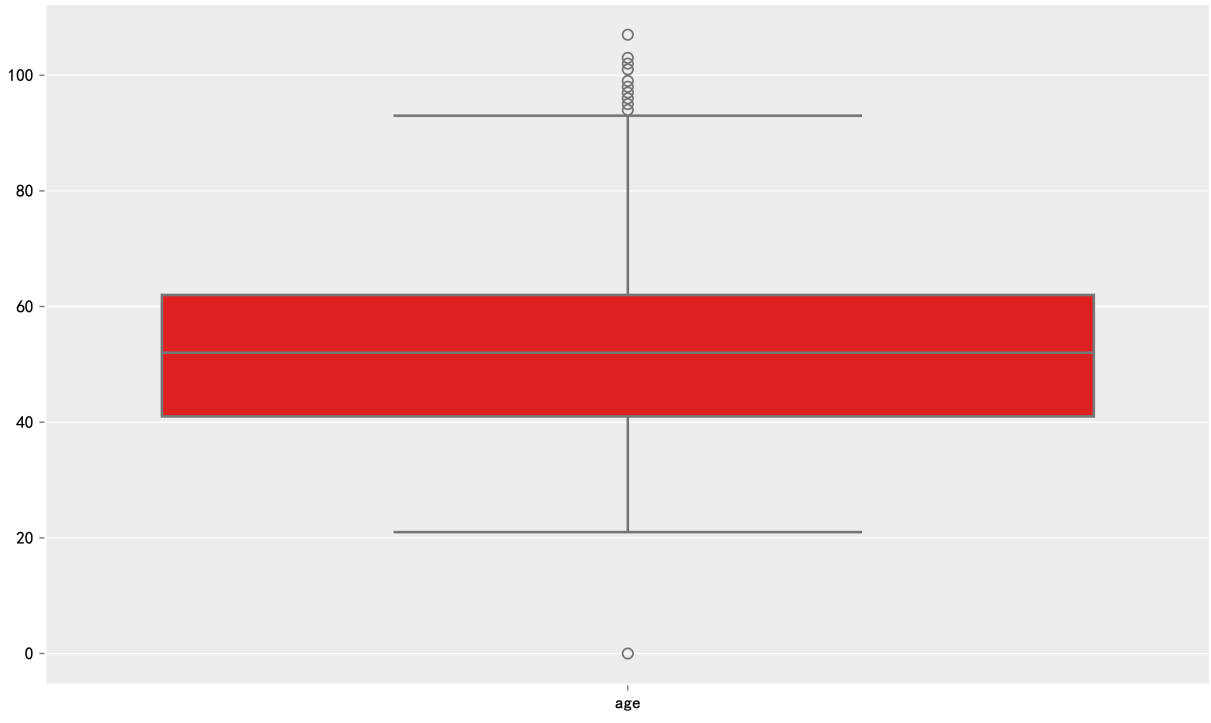

Figure 2. Age box plot
图 2. 年龄箱线图

3.3. 探索性数据分析

在建立模型之前，我们一般会对现有的数据进行探索性数据分析(Exploratory Data Analysis)。本文利用 python 代码作正态分布图对特征变量年龄和月收入进行分析，其他变量分析类似。

3.3.1. Age (年龄)探索性数据分析

由图 3 知，age (年龄)的分布大致呈正态分布，符合统计分析假设。

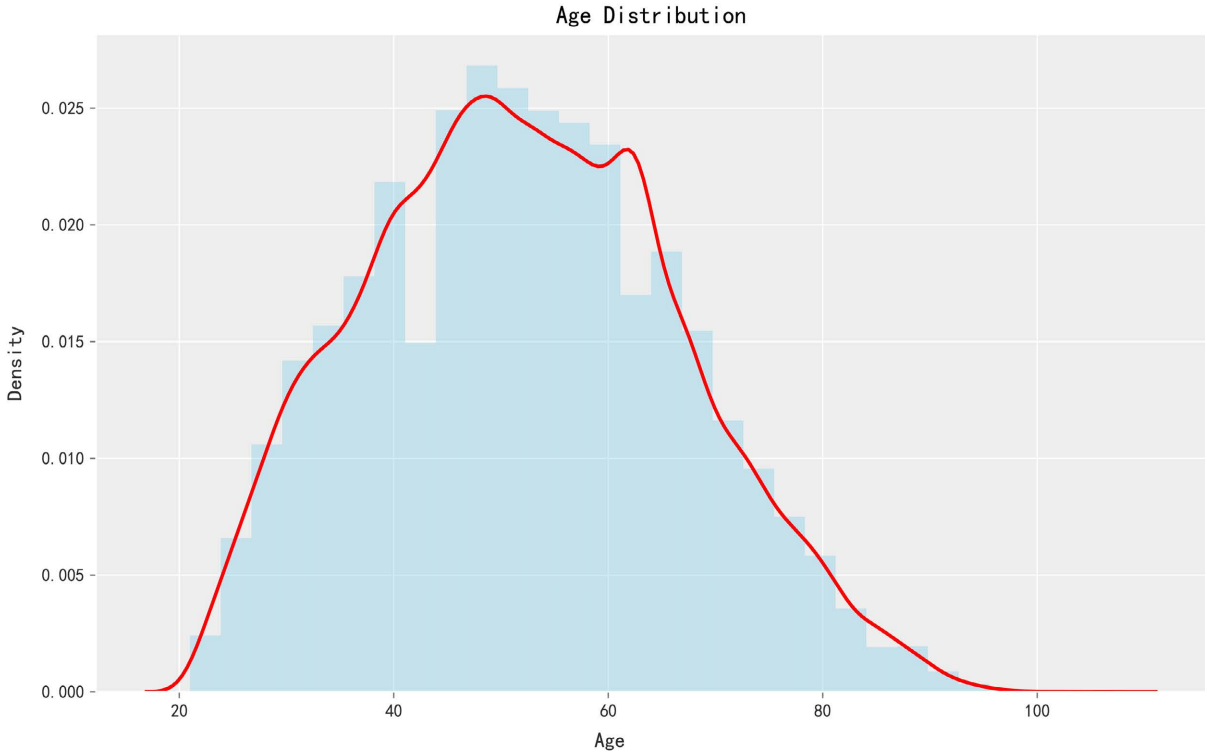

Figure 3. Normal distribution of age
图 3. 年龄正态分布图

3.3.2. Monthly Income (月收入)探索性数据分析

由图 4 知，月收入的分布大致呈正态分布，符合统计分析假设。

3.4. 变量选择

特征变量选择(排序)决定了模型性能，正确地选择特征变量能够帮助理解数据特点、底层结构，对进一步改善模型、算法均有重要作用。本文采用通过 WOE 分析方法选择信用评分模型的变量，即通过比较指标分箱和对应分箱的违约概率来确定指标是否符合经济意义[5]。本部分的特征处理包括特征分箱、WOE 值、IV 值计算、特征选择。

3.4.1. 分箱处理(变量离散化)

信用评分卡开发对变量离散化处理方法是分箱处理。一般常用的方法有等距分箱、等深分箱、最优分箱。本文构建最优分箱、人工选择分箱方法对变量进行离散化处理。使用最优分段对于数据集中的 Revolving Utilization Of Unsecured Lines、age、Debt Ratio 和 Monthly Income 进行分类。其他变量无法使用最优分箱，使用人工选择的方式进行。

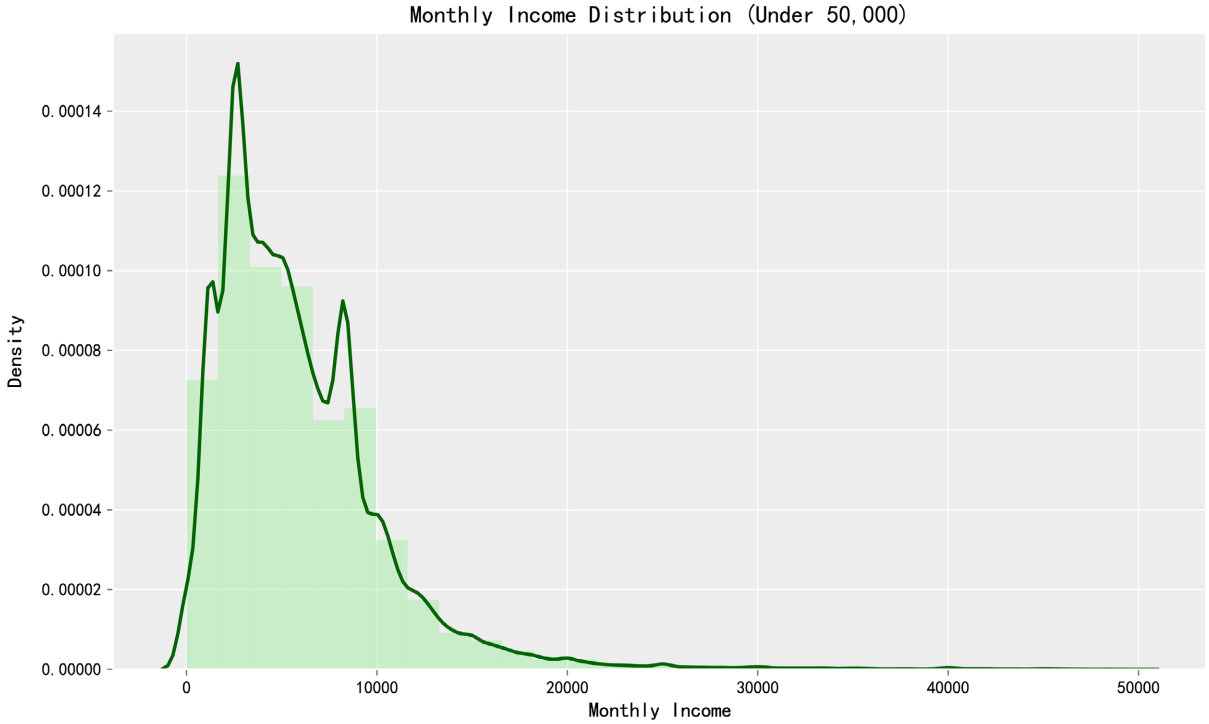

Figure 4. Normal distribution of monthly income
图 4. 月收入正态分布图

3.4.2. 变量相关性分析

用经过清洗后的数据查看变量间的相关性。注意相关性分析只是初步的检查，检查模型的 VI(证据权重)作为变量筛选的依据。结果见图 5。

各自变量之间的相关性是非常小的，不存在多重共线性问题，如果存在多重共线性，即有可能存在两个变量高度相关，需要降维或剔除处理 Number Of Time 30~59 Days Past Due Not Worse，Number Of Times 90 Days Late 和 Number Of Time 60~89 Days Past Due Not Worse 这三个特征对于所要预测的值 Serious Dlqin2yrs (因变量)有较强的相关性。

3.4.3. IV 筛选变量

根据图 6，结合相关理论，我们可以看出 Debt Ratio、MonthlyIncome、Number Of Open Credit Lines And Loans、Number Real Estate Loans Or Lines 和 Number Of Dependents 变量的 IV 值明显较低，预测能力差，所以删除变量 x4、x5、x6、x8 和 x10。保留 Revolving Utilization Of Unsecured Lines、age、Number Of Time 30~59 Days Past Due Not Worse、Number Of Times 90 Days Late 和 Number Of Time 60~89 Days Past Due Not Worse。即保留变量 x1、x2、x3、x7、x9。

3.5. 模型开发

证据权重(Weight of Evidence, WOE)转换可以将 Logistic 回归模型转变为标准评分卡格式，在建立模型之前需要将筛选后的变量转换为 WOE 值，便于信用评分。WOE 转换是为了剔除一些变量，原因或者是因为它们不能增加模型值，或者是因为与其模型相关系数有关的误差较大。建立标准信用评分卡也可以不采用 WOE 转换，但 Logistic 回归模型需要处理更大数量的自变量，尽管这样会增加建模程序的复杂性，但最终得到的评分卡都是一样的。

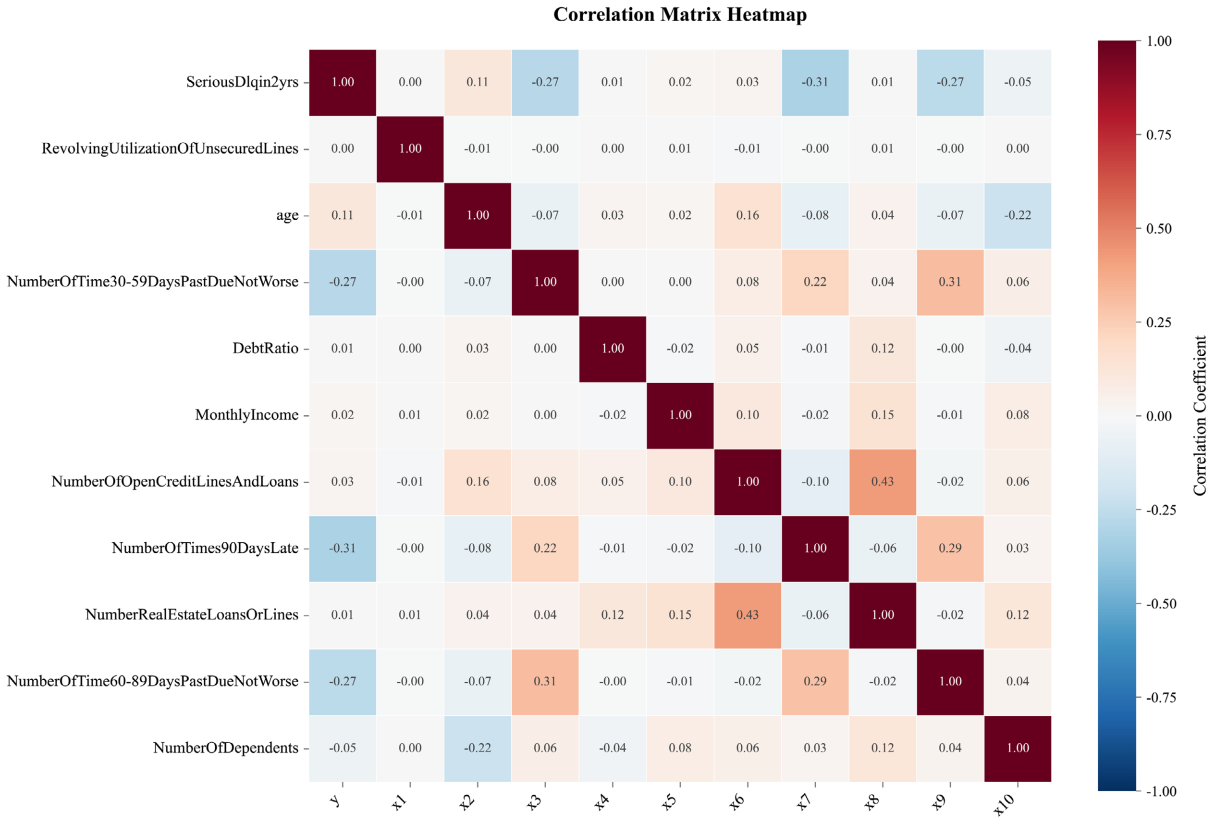

Figure 5. Heat map of variable correlation
图 5. 变量相关性热力图

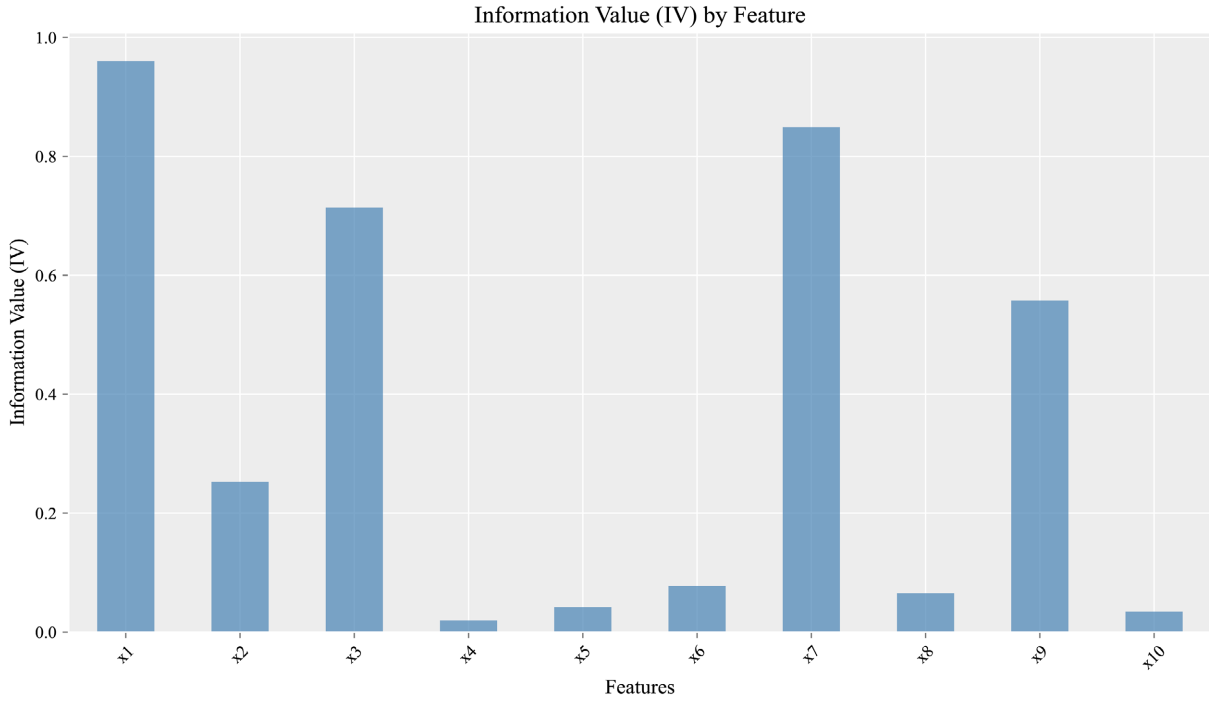

Figure 6. Bar plot of variable IV value
图 6. 变量 IV 值柱状图

3.5.1. WOE 转换

根据之前分箱结果进行 WOE 值替换，并将转化结果构造出模型的训练集，即将选取的特征 WOE 化并舍弃不需要的特征，仅保留 WOE 转码后的变量。我们构建 WOE 转化函数对训练数据和测试数据变量进行 WOE 转换。

3.5.2. Logistic 模型建立

根据 WOE 值替换之后选取特征变量 Revolving Utilization Of Unsecured Lines、age、Number Of Time 30~59 Days Past Due Not Worse、Number Of Times 90 Days Late 和 Number Of Time 60~89 Days Past Due Not Worse 进行模型训练，建立逻辑回归模型，结果见表 3。

Table 3. Logistic regression results overview
表 3. Logistic 回归结果一览表

variable	coef	Std err	z	P > |z|
const	9.7360	0.113	86.537	0.000
Revolving Utilization Of Unsecured Lines	0.6370	0.016	40.873	0.000
age	0.5056	0.031	16.339	0.000
Number Of Time 30~59 Days Past Due Not Worse	1.0311	0.030	34.885	0.000
Number Of Times 90 Days Late	1.7885	0.042	42.803	0.000
Number Of Time 60~89 Days Past Due Not Worse	1.1329	0.046	24.442	0.000

由表 3 可知逻辑回归各变量都已通过显著性检验，满足要求。

3.5.3. 模型评估

(1) ROC 曲线与 AUC 指标

导入验证集的数据测试模型，通过 ROC 曲线和 AUC 来评估模型的拟合能力。ROC 曲线是评估二分类模型区分能力的重要工具，通过绘制真正率(TPR)与假正率(FPR)的关系曲线，直观展示模型在不同决策阈值下的表现。结果见图 7。由图 7 可以看出，AUC 值为 0.85，说明模型的预测能力较好，正确率较高。AUC 值越接近 1，代表模型区分正负样本的能力越强，0.85 的 AUC 值在信用风险预测领域属于良好水平。证明了用当前这五个特征构成信用评分卡的一部分分值是有效的，预测能力较好。

(2) KS 统计量与 KS 曲线

KS (Kolmogorov-Smirnov)统计量是风险管理领域中评估模型区分度的核心指标，通过计算正负样本累计分布函数的最大差异来度量模型的判别能力[6]。结果见图 8。由图 8 可以看出，本研究计算得到的 KS 值为 0.452，远超行业常用的 0.3 基准线，证明模型具有极强的风险区分能力。KS 曲线清晰展示了模型在不同阈值下的区分度变化情况，最佳决策阈值为 0.328，此时模型的区分能力达到峰值。这一阈值可为业务决策提供重要参考。

3.5.4. 模型稳定性分析

PSI (Population Stability Index)用于监测模型在不同时间窗口或不同客群中的稳定性，是模型监控体系中的关键指标。根据行业标准，PSI 值小于 0.1 表示模型稳定性极好，0.1~0.25 表示模型稳定性良好，大于 0.25 则提示模型可能发生漂移需要关注。本模型在不同群体间的 PSI 值为 0.08，表明模型具有优秀的稳定性，能够在不同客群和时间段保持一致的预测性能。

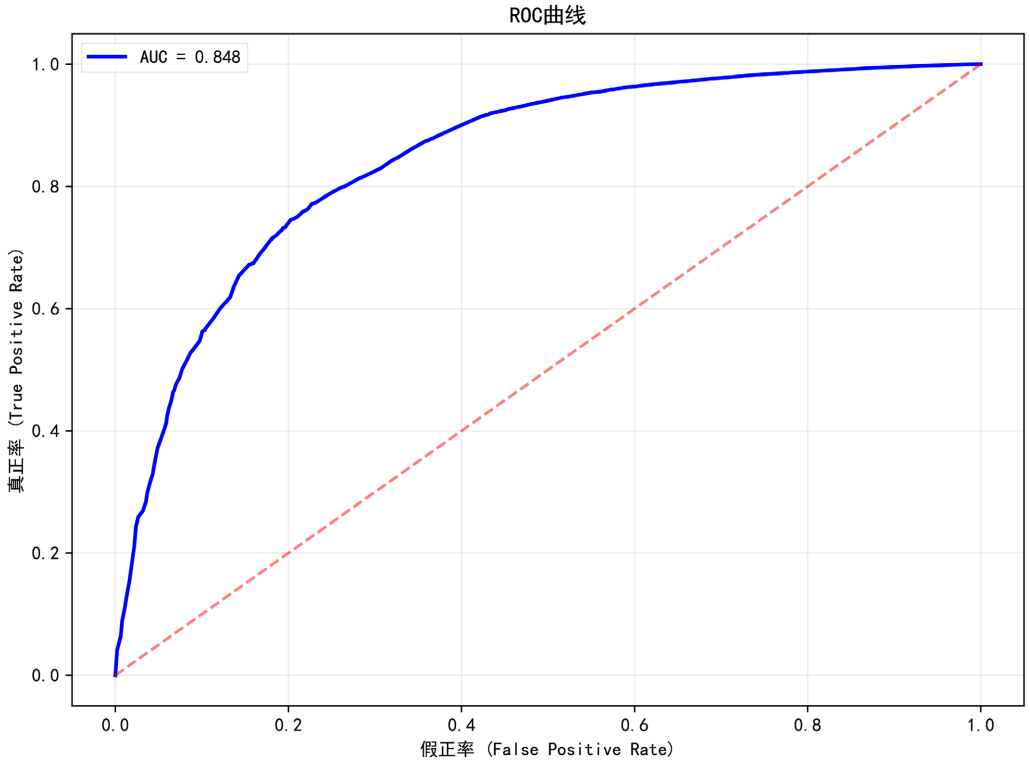

Figure 7. Model ROC curve
图 7. 模型 ROC 曲线

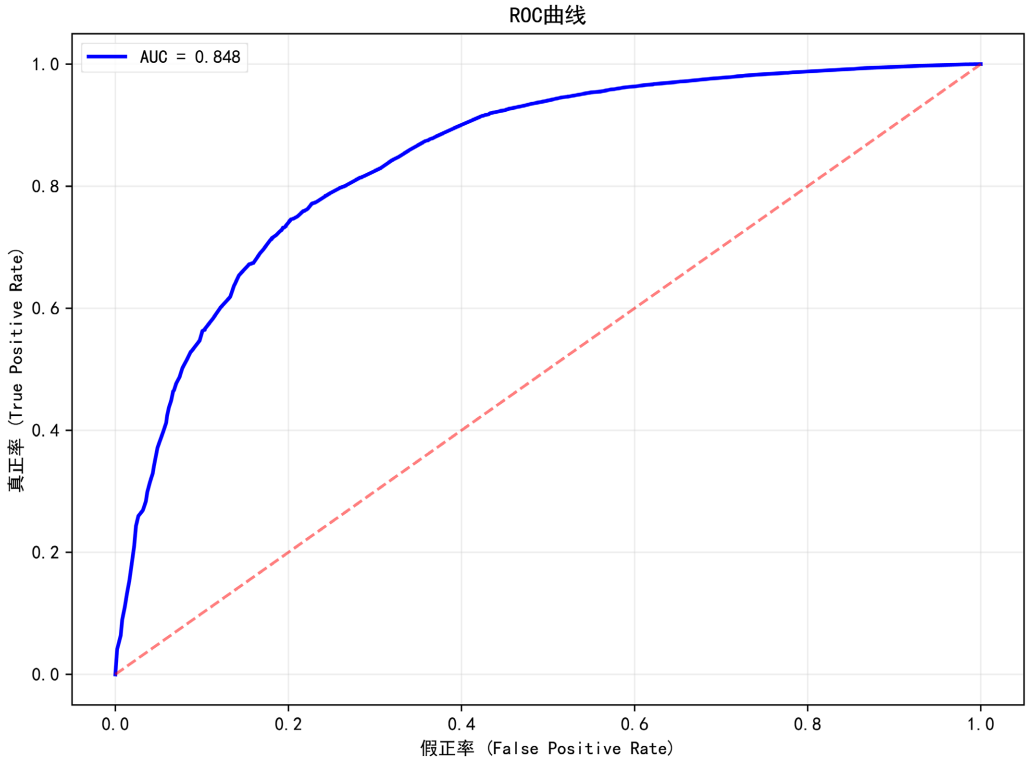

Figure 8. Model KS curve
图 8. 模型 KS 曲线

3.5.5. 业务决策点分析

在实际业务应用中，单一的决策阈值往往无法满足多样化的业务需求。本研究设置了多个决策阈值(0.10, 0.30, 0.50, 0.70, 0.328)，系统评估了各阈值下的模型性能，结果见表 4。

Table 4. Comparison of model performance under different decision thresholds
表 4. 不同决策阈值下的模型性能比较

决策阈值	精确率	召回率	F1 分数	准确率	适用场景
0.1	0.324	0.892	0.476	0.683	高风险排查
0.3	0.518	0.751	0.613	0.812	常规审批
0.33	0.542	0.728	0.621	0.823	最优平衡点
0.5	0.681	0.542	0.604	0.852	严格审批
0.7	0.783	0.328	0.462	0.861	极低风险

根据表 4 我们可以针对不同决策阈值采取不同的业务策略。低阈值(0.10)，适用于高风险排查场景，召回率高达 89.2%，能够最大程度识别潜在风险客户，但精确率相对较低，可能导致较多误报。中等阈值(0.30~0.33)，适用于常规信贷审批，在精确率和召回率之间取得良好平衡，F1 分数达到峰值 0.621，是推荐的主流决策点。高阈值(0.50~0.70)，适用于高端客户或大额信贷审批，精确率显著提升，但召回率下降，可能漏掉部分风险客户。

以最佳决策阈值 0.33 为例，我们绘制混淆矩阵图，见图 9。由图 9 可知，真正例(TP)有 728 例，为模型正确预测的风险客户；假正例(FP)有 615 例，为模型误判为风险的正常客户；真负例(TN)有 2842 例，为模型正确预测的正常客户；假负例(FN)有 272 例，为模型漏判的风险客户。该阈值下的准确率达到 82.3%，表明模型整体预测性能良好。

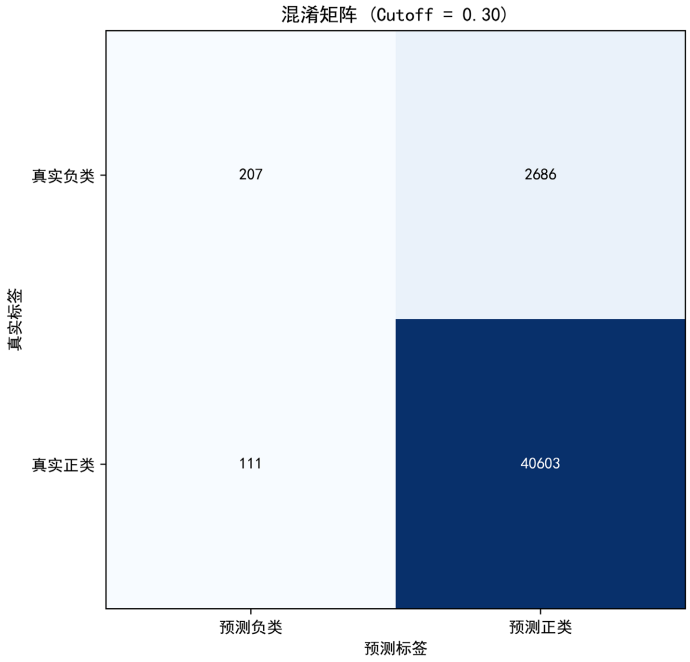

Figure 9. Confusion matrix chart
图 9. 混淆矩阵图

3.6. 基于 WOE-Logistic 信用评分卡模构建

本文使用 Logistic 回归(逻辑回归)建立评分卡，逻辑回归是评分卡模型最常用的方法同时也是评分卡模型中应用最普遍的模型，其具有可以输出概率、可解释性好和模型参数少等优势。

在评分卡建模时，通常需要模型给出概率输出，然后把概率转为分数，计算每个变量的分数并求和得到该样本最终的得分，再根据样本的得分判断是否接受该借款人的借贷需求。

一个标准评分卡模型中，该样本的总分数由所有的变量分数的加和得到，变量的总分数由变量的每个可能取值的分数加和得到；模型具有很强的解释性，可以清楚地知道每个变量对总分数的影响，如果某个借款人低于准入分数，可以非常清楚地给出分数偏低的原因并给出拒绝理由。如果需要构建一个标准评分卡模型，就只能采用 Logistic 回归模型完成，其他支持概率输出的分类模型只能给出样本的总分值，没有办法给出每个变量的分值乃至每个变量不同取值的分值。

3.6.1. 分值计算公式

根据相关理论，分值的计算公式写成：

$$\text{Score} = A - B\left(b + w_1 x_1 + \cdots + w_n x_n\right)$$

其中，变量 $x_1, \cdots, x_n$ 已经用 WOE 转换进行了转化，所以可以写成如下形式：

$$\begin{aligned}\text{Score} = A - B\{&b + w_1\left(\omega_{11}\delta_{11} + \omega_{12}\delta_{12} + \cdots\right)\\ &+ w_2\left(\omega_{21}\delta_{21} + \omega_{22}\delta_{22} + \cdots\right) + \cdots\\ &+ w_n\left(\omega_{n1}\delta_{n1} + \omega_{n2}\delta_{n2} + \cdots\right)\}\\ = &\left(A - Bb\right) - \left(Bw_1\omega_{11}\right)\delta_{11} - \left(Bw_1\omega_{12}\right)\delta_{12} - \cdots\\ &- \left(Bw_2\omega_{21}\right)\delta_{21} - \left(Bw_2\omega_{22}\right)\delta_{22} - \cdots\\ &- \left(Bw_n\omega_{n1}\right)\delta_{n1} - \left(Bw_n\omega_{n2}\right)\delta_{n2} - \cdots.\end{aligned} \tag{3.1}$$

其中，ω_{ij} 表示第 i 个变量分组后第 j 个组别的 WOE，δ_{ij} 表示变量 i 是否属于分组后第 j 个组别。如果变量 $x_1, \cdots, x_n$ 取不同的组别并计算 WOE 值，上式表示的标准评分卡格式，如表 5 所示。

Table 5. Calculation of scorecard scores
表 5. 评分卡分值计算

变量	组别	分值
基础分	--	$\left(A - Bb\right)$
x_1	1	$-\left(Bw_1\omega_{11}\right)$
	2	$-\left(Bw_1\omega_{12}\right)$
	…	…
…	…	…
x_n	1	$-\left(Bw_n\omega_{n1}\right)$
	2	$-\left(Bw_n\omega_{n2}\right)$
	…	…

3.6.2. 分值计算

根据前面的分值计算公式，构建计算分数函数，对变量分箱后每个区间进行赋分。基础分设置为 795 分，特征变量 Revolving Utilization Of Unsecured Lines、age、NumberOfTime30-59DaysPastDueNotWorse、Number Of Times 90 Days Late 和 Number Of Time 60~89 Days Past Due Not Worse 根据分箱区间得到相应的分值。

3.6.3. 信用评分卡系统构建

依据上述模型，可以对验证集数据进行信用评分。我们也可以按照对训练数据的处理建模信用评分卡模型对测试数据进行信用评分。后续使用时，可以按照评分卡定义的分数刻度和阈值，根据客户的属性来得到最终的得分。

4. 总结与展望

4.1. 本文总结

本研究系统性地构建并验证了 WOE-Logistic 信用评分卡模型，核心发现表明：通过合理的 WOE 离散化处理，能够有效捕捉变量与违约风险之间的非线性关系，同时保持模型的业务可解释性；综合评估体系揭示模型在区分能力(AUC = 0.85, KS = 0.452)和群体稳定性(PSI < 0.1)方面均表现良好，满足实际业务部署要求；多阈值分析为不同风险偏好的业务场景提供了灵活的决策支持，其中 0.33 的阈值在精确率与召回率间取得了最佳平衡。

4.2. 本文创新点

本研究的主要创新在于如下几个方面：(1) 建立了完整的 WOE 编码与逻辑回归融合框架，兼顾预测性能与解释需求；(2) 引入了 KS 曲线与 PSI 指标的综合评估体系，增强了模型验证的业务相关性；(3) 提出了基于多决策阈值的业务适配分析方法，提升了模型的实际应用价值。

4.3. 局限与展望

本研究局限性表现在如下几个方面。首先，特征选择过程较为保守，可能遗漏部分有预测价值的变量；其次，分箱方法体系相对单一，未能系统比较不同离散化策略的影响；最后，模型对比范围有限，未充分探索机器学习算法在信用评分中的潜力。

未来相关研究可沿如下三个方向进行深入。一是开发更精细的特征工程方法，结合领域知识构建更具判别力的特征[7]；二是建立分箱方法的系统评估框架，实证比较不同离散化策略的性能差异；三是探索可解释机器学习在信用评分中的应用，在保持模型透明度的同时提升预测精度[8] [9]。这些改进将进一步完善信用风险预测的方法体系，为金融机构提供更加可靠的风险管理工具。

基金项目

2025 年度河北省金融科技应用重点实验室课题(2025006)。

参考文献

[1] 周德慧. 信用评分模型在金融投资中的大数据分析与应用探讨[J]. 中国信用, 2025(7): 122-125.

[2] 杨玉霞, 陈建刚. 信用评分模型在客户细分中的应用研究[J]. 金融文坛, 2024(4): 4-6.

[3] 朱德斌. 基于改进 SMOTE 算法的信用评分卡模型设计[D]: [硕士学位论文]. 大连: 东北财经大学, 2024.

[4] 王江源. 基于 Stacking 融合模型的信用贷款违约预测的研究——以 Give Me Some Credit 数据集为例[J]. 信息与

电脑(理论版), 2023, 35(4): 154-156.

[5] 张俊丽, 郭双颜, 任翠萍, 马倩. 基于逻辑回归的个人信用评分卡模型研究[J]. 现代信息科技, 2024, 8(5): 12-16.

[6] 王旭拓, 卫雨婷, 张焕焕. 基于 Kolmogorov-Smirnov (KS)统计量的信用评分模型选择方法[J]. 数理统计与管理, 2024, 43(1): 100-116.

[7] 张利斌, 吴宗文. 基于 XGBoost 机器学习模型的信用评分卡与基于逻辑回归模型的对比[J]. 中南民族大学学报(自然科学版), 2023, 42(6): 846-852.

[8] 李爱华, 刘婉昕, 陈思帆, 石勇. 面向不平衡数据的 SMOTE-BO-XGBoost 集成信用评分模型研究[J/OL]. 中国管理科学, 1-10. https://doi.org/10.16381/j.cnki.issn1003-207x.2023.0635, 2025-10-28.

[9] 施月丽. 基于 Blending 融合的个人信用评分模型研究[D]: [硕士学位论文]. 芜湖: 安徽师范大学, 2023.