

# 基于元学习的财务舞弊识别模型

张学勇 施懿

中央财经大学金融学院，北京

**摘要：**本文基于多元化数据挖掘和机器学习集成方法改进两个方面，对于如何识别财务舞弊并提高识别效率问题提供了系统性预测方法。在多元化数据方面，不仅对传统财务因子进行了重构，而且引入公司治理层面因子并利用文本分析构建了语言类因子。在机器学习集成方法改进方面，以9种不同特质的机器学习算法作为基学习器，套用元学习框架对上市公司财务舞弊进行系统性识别。本文发现，（1）元学习框架能够显著提升舞弊样本召回率和预测精确度，提高整体学习器预测性能，并且对于大部分行业都有效果。（2）接近真实场景的滚动预测方法下，元学习框架依然能显著提高基学习器的财务舞弊识别能力。（3）公司治理因子、语言类因子对于财务舞弊识别有一定的帮助。

**关键词：**财务舞弊 元学习 机器学习 文本分析

中图分类号： F832 文献标识码： A

---

本论文感谢国家自然科学基金（71773152，71403306，71673318，71602198）资助

通讯作者：张学勇中央财经大学金融学院 北京 100081，[zhangxueyong@cufe.edu.cn](mailto:zhangxueyong@cufe.edu.cn)

邮寄地址：北京海淀区学院南路中央财经大学研究生院 100081，13701249262

作者简介：

张学勇，中央财经大学金融学院教授、博士生导师，中央财经大学研究生院院长，主要研究方向是：机器学习与公司财务。联系方式：13701249262，[zhangxueyong@cufe.edu.cn](mailto:zhangxueyong@cufe.edu.cn)

施懿，中央财经大学金融学院研究生。联系方式：13051106816，[cufe\\_sy2018@163.com](mailto:cufe_sy2018@163.com)

# 基于元学习的财务舞弊识别模型

张学勇 施懿

中央财经大学金融学院，北京

**摘要：**本文基于多元化数据挖掘和机器学习集成方法改进两个方面，对于如何识别财务舞弊并提高识别效率问题提供了系统性预测方法。在多元化数据方面，不仅对传统财务因子进行了重构，而且引入公司治理层面因子并利用文本分析构建了语言类因子。在机器学习集成方法改进方面，以9种不同特质的机器学习算法作为基学习器，套用元学习框架对上市公司财务舞弊进行系统性识别。本文发现，（1）元学习框架能够显著提升舞弊样本召回率和预测精确度，提高整体学习器预测性能，并且对于大部分行业都有效果。（2）接近真实场景的滚动预测方法下，元学习框架依然能显著提高基学习器的财务舞弊识别能力。（3）公司治理因子、语言类因子对于财务舞弊识别有一定的帮助。

**关键词：**财务舞弊 元学习 机器学习 文本分析

中图分类号： F832 文献标识码： A

## Financial Fraud Recognition Model Based on Meta-learning

Xueyong Zhang, Yi Shi

School of Finance, Central University of Finance and Economics

**Abstract:** Based on diversified data-mining and machine learning integration, this paper provides a systematic prediction method for how to identify financial fraud and how to improve the efficiency of financial fraud identification. In terms of diversified data-mining, not only the traditional financial factors are reconstructed, but also the corporate governance and the language factors are used. For improving the machine learning efficiency, nine machine learning algorithms with different characteristics are used as the basic learners and the meta-learning framework is applied to systematically identify the financial fraud of listed companies. This paper finds that: (1) Meta-learning framework can significantly improve the recall and precision of fraud samples, the overall prediction performance of the learner, and take effect on most industries. (2) By rolling prediction method close to the real scenario, the meta-learning can still significantly improve the financial fraud identification ability of the basic learner. (3) corporate governance factors and language factors are helpful for financial fraud identification.

**Key Words:** Financial fraud; Meta-learning; Machine Learning; Text analysis

## 一、引言

财务舞弊一直困扰着资本市场建设并愈发成为一个严重的问题（Throckmorton et al., 2015; Albashrawi and Lowell, 2016; Hajek and Henriques, 2017），也引起了学术界的重视（Spathis, 2002; Lin et al., 2015）。过去的经验表明，上市公司财务舞弊对投资者乃至整个资本市场带来了严重的负面影响。Beasley 等（2010）统计发现，在初次被爆财务造假的公司的事后股价平均下跌 16.7%，并且有 47% 的财务造假公司退市。另外，美国历史上十大破产案中有四起也与重大财务欺诈相关（Abbasi, 2012）。近年来，财务舞弊事件无论从发生频率还是涉及金额都呈现上涨趋势，愈发成为投资者、审计师、监管方所担忧的问题（Hajek and Henriques, 2017）。财务舞弊不仅摧毁了投资者对于财务报表质量的信任，而且还严重扰乱了资本市场的正常运转。因此，如何提高预测上市公司财务舞弊效率，这一命题在当前环境下的重要性也愈发凸显出来。

尽管财务舞弊识别需要具备丰富的经验知识以及依赖外部审计给予客观的审计意见，但从过去曝光的案例来看，传统财务舞弊手段由于依靠人工、效率较低，并不能较好地适应当前财务舞弊识别的需求（Dyck et al., 2010; Hajek and Henriques, 2017）。首先，舞弊手段逐渐变得繁复，舞弊过程不再是单一类型舞弊，可能同时涉及利润表、资产负债表、虚假披露舞弊等，财务舞弊识别也变得更为复杂且耗时耗资（朱锦余、高善生, 2007; West and Bhattacharya, 2016）。其次，从财务舞弊披露时点来看，财务舞弊处罚具有延迟性。根据国泰安数据统计，2006 年-2018 年上市公司财务舞弊处罚案例中，延迟 1-4 年占所有案例 64.15%，17.18% 的公司延迟了 5 年及以上才被披露其舞弊行为，这反映了财务舞弊具有隐蔽性，导致财务舞弊识别不及时，放大了财务舞弊的恶性影响。

从过去的研究来看，大部分研究都是从财务舞弊识别因子和舞弊识别方法两个角度出发构建财务舞弊识别模型，试图以模型判断来代替人工识别。从财务舞弊识别因子构建角度，除了传统财务指标外，研究还囊括了公司治理和文本挖掘两方面的指标，公司治理维度包括高管人员背景特征、董事会特征、股权结构等（Chen et al., 2006; Hasnan et al., 2012），文本挖掘维度包括媒体关注的频率和报道基调（吴芃等, 2019）、财务报表附注和管理层讨论与分析（MD&A）文本的分析（Newman et al., 2003; Throckmorton et al., 2015; Purda and Skillicorn, 2015; Dong et al., 2016）等。但较少有学者关注如何将财务因子、公司治理因子以及语言因子进行结合，从而提升财务舞弊识别的准确率（Hajek and Henriques, 2017）。从舞弊识别方法角度，近年来随着计算机技术的提高以及机器学习模型的研究方法不断深入，财务舞弊识别的准确率也将逐步提高。但目前财务舞弊识别样本存在数据不平衡的问题，一些学者仍采用 1:1 或者固定舞弊与非舞弊样本进行

模型训练，不符合真实应用场景。并且目前选取模型的方法主要通过后验结果导向选取，而不同时期可能适用于不同的模型（Brazdil et al., 2009），这种主观后验选取的方法不能适应新的数据特征分布。

因此，如何提升财务舞弊识别效率并同时适应不同的数据分布特征是本文的研究目标，而本文也将运用前沿的机器学习算法与元学习框架，将文本词性分析与金融市场相结合，探索机器学习与大数据工具在金融领域的应用。本文的贡献在于：（1）结合财务因子、公司治理因子、语言因子，丰富了我国财务舞弊识别的因子库，也为财务报表文本分析应用提供了新的实证支持；（2）套用元学习框架，提供了一套完整的提高财务舞弊识别准确率的方法，包括解决数据不平衡问题、因子筛选、模型调参以及模型预测，为提升财务舞弊识别预测准确率提供了可实行的客观方案。

本文后续内容主要包括以下部分：第二部分回顾已有研究，第三部分对模型流程进行介绍，包括因子筛选、数据不平衡处理以及元学习框架预测流程，其中，Logistic 回归、朴素贝叶斯、支持向量机、K 近邻算法、决策树、随机森林、梯度提升树、极限梯度提升树、神经网络作为元学习框架中的基学习器。第四部分介绍数据来源和因子构建，第五部分实证部分分析了哪些因子对于财务舞弊识别有帮助，并从不同角度展示元学习框架在财务舞弊识别模型中的预测性能。本文模型具体流程如下图：

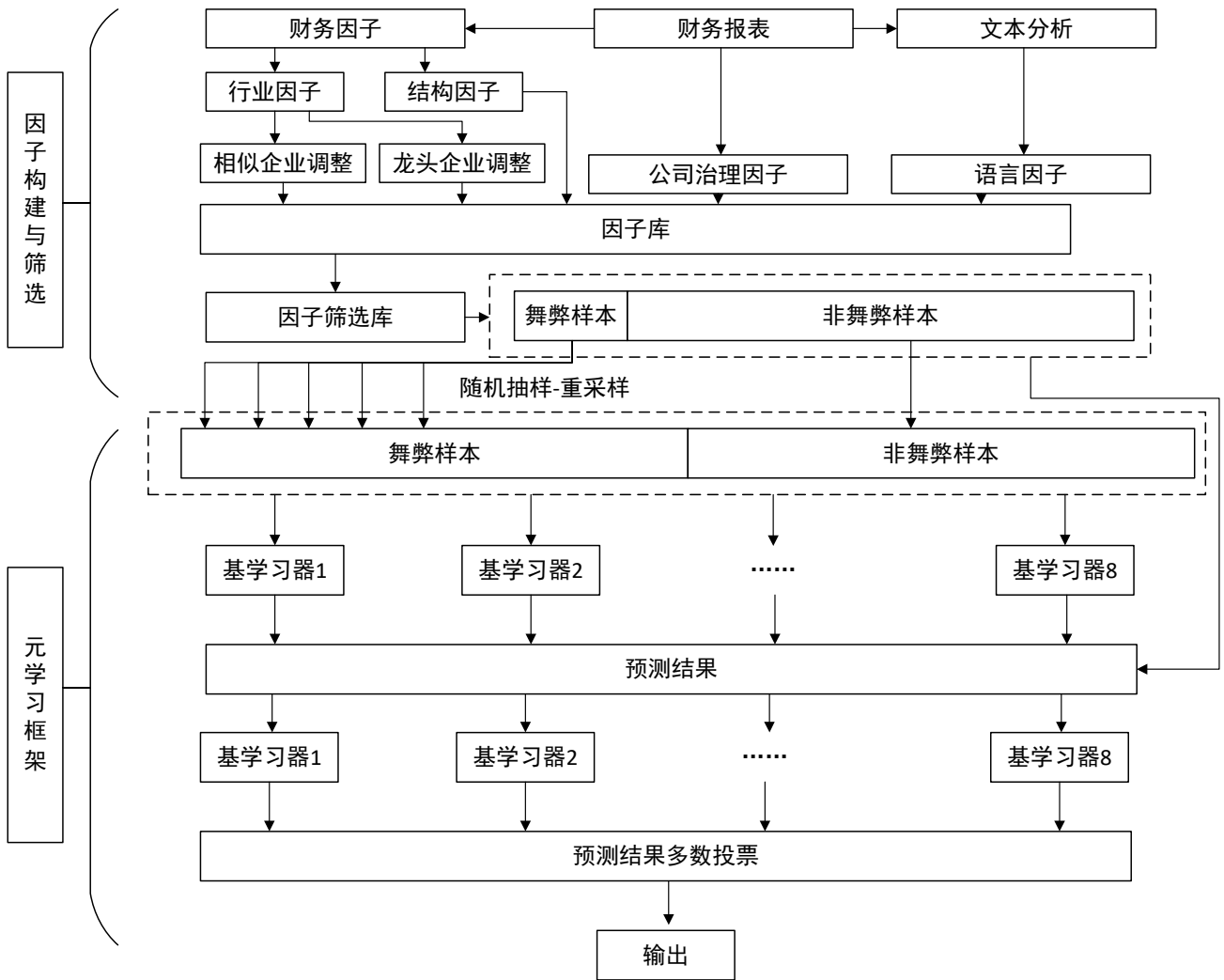


图 1 模型流程图

## 二、文献综述

上市公司财务舞弊的识别不仅是投资者、审计师、监管方所担忧的问题，也是当前资本市场和国内外学术界关注的热点问题之一。

传统财务识别模型基于财务舞弊动因理论而发展起来，其中应用最为广泛的是舞弊三角理论和 GONE 理论。舞弊三角理论将动机或压力、机会、态度或借口作为识别舞弊的重要条件，其中，实施舞弊的动机或压力是舞弊发生的首要条件。GONE 理论由 Bologna et al.(1993)提出，文章认为企业舞弊行为由 G(Greed)、O (Opportunity)、N (Need)、E (Exposure) 四个影响因素驱动导致，之后他们在此基础上不断归纳，最终形成舞弊风险因子理论。

美国注册会计师协会分别在 1988 年、1997 年、2002 年发布了审计准则（SAS），明确规定了审计师审查财务舞弊的规范，因此，也开始有学者基于 SAS 所提及的风险因子构建了财务舞弊识别模型。一方面，一些学者采用问卷、清单的形式获得 SAS 准则中提及的一些风险因子（Pincus, 1989; Beasley, 1996），但这些因子的可靠性无法确认（Wilks and Zimbelman, 2004）。另一方面，SAS 几乎没有提供如何利用风险信号或者风险因子来进行判断，因此审计师还需要其他辅助工具进行判断，因此，整个审计流程是非定量且非自动化的（Song et al., 2014）。

还有一些学者用公开财务数据对财务舞弊风险进行建模。国外较早的系统性识别财务舞弊模型是 M-score 模型，通过财务指标利用历史数据构建线性方程（Beneish, 1999）。之后在该基础上丰富了财务舞弊风险因子构成了 F-score 模型（Bell and Carcello, 2000），其中 F-score 预测准确率已达 69%。但这些模型仅涉及财务因子，事实证明公司治理、文本因子等非财务因子对于财务舞弊识别也有一定帮助（胡奕明和唐松莲, 2008; 李丹和宋衍蘅, 2010; Goel and Uzuner, 2016; Petr and Roberto, 2017），并且随着财务舞弊风险因子增多、模型维数上升，大量噪音、不相关的、冗余特征充斥在数据中（Seo and Choi, 2016），模型估计准确度也随之下降。

随着机器学习算法研究的深入和方法论的丰富，一些学者将机器学习算法应用到财务舞弊识别中，并且相较于传统财务识别模型有着明显的优势。首先，机器学习擅长挖掘数据之间的非线性特征（Hastie et al., 2009），并且准确率得到提升。传统 M-score、F-score 模型仅能证明财务因子对于识别财务舞弊具有一定的解释力，而缺乏实际证据支持其预测准确效果。而已有学者证明，机器学习准确率基本在 80% 以上（Ravisankar et al., 2011; Lin et al., 2015; Hajek and Henriques, 2017）。并且，这些不同算法中包含的激活函数、核函数形式都丰富了拟合函数以便适应预测目标的复杂性。

其次，机器学习可以采用集成模式而非单一模型。传统模型通常采用单个模型进行预测，而机器学习不仅算法多样化，而且同一算法中根据不同的参数组合训练出的模型也具有多样性。对于基础机器学习算法的优化也可以通过结合模型（Combination）或者挑选适合的模型来提高预测模型的准确度（Vilalta and Drissi, 2002; Cui et al., 2016）。集成学习的思路是通过合并多个弱学习器，提升机器学习性能以获得更好的预测结果。集成学习分类三类，一是用于减少偏差的 boosting，二是用于减少方差的 bagging，三是用于提升预测结果的 stacking。元学习是属于 stacking，从机器学习或者数据挖掘中得到数据，用于提升预测结果的质量。通常，机器学习多种算法提供了解决问题的一系列方法，但是并没有告诉我们，在给定背景下哪些算法更适合使用，元学习则提供了一种方式能够学习到在学习过程中哪一种算法和潜在特征能够被

更有效的运用（Brazdil et al., 2009; Lemke et al., 2015），以解决复杂而又动态的财务舞弊识别问题，并且随着时间推移而增强学习能力（Matijaš et al., 2013）。

第三，机器学习算法为如何解决数据不平衡问题提供了客观方案。传统模型一般采用匹配样本进行解决，但匹配近似样本的标准较为主观，并且数据样本与真实世界不符（Hoogs et al., 2007; Sagadevan et al., 2018）。已有学者采用过采样的方式，或者欠采样生成多个子数据集，分别训练不同的模型，再将模型进行集成输出（Glancy and Yadav, 2011），这样的好处是利用现有全部数据，并且避免主观挑选匹配样本，导致训练样本与实际样本分布偏差过大，避免出现过拟合的情况。

另外，除了财务舞弊模型方法改进之外，非金融变量也开始进入学术界的视野。公司治理层面因子也有助于量化财务舞弊机会，包括高管层持股情况、高管个人背景、股权结构等（Lin et al., 2015）。另外，文本分析生成的语言类因子也成为学术界识别财务舞弊的新工具。Humpherys 等（2011）通过词汇多样性和句法复杂性等语言因子获得了接近 70% 的预测准确率。Goel 和 Uzuner（2016）则发现财务造假的公司会同时更多的使用积极和消极的词汇，而不是采用中性词。Dong 等（2016）将文本因子分类为主体、观点、情感、情态、人称代词、写作风格、题材七个类型，发现加入语言因子的模型平均准确率高达 82.36%，显著优于仅采用财务比率的基准方法。Petr 和 Roberto（2017）则得出使用否定词较少的公司存在欺诈行为的可能性较低的结论。这也说明了数据挖掘新因子对于识别财务舞弊的重要性。

### 三、模型流程

本文采用的基学习器包括决策树、梯度提升树（GBDT）、K 近邻算法、Logistic 回归、朴素贝叶斯算法、随机森林算法、支持向量机、极限梯度提升树（XGBoost）、神经网络（Neural Network）。本文主要用年报数据与季度报告数据作为基学习器的数据源，基学习器输出的预测值作为下一层堆叠分类器的输入源，即从底层数据输入、基学习器训练，再到堆叠训练整个过程称为元学习。

#### （一）数据处理

本文先对全部因子作为全部解释变量进行模型训练，以不同学习器在测试集表现的多数投票结果作为基准。其次，用信息增益率（IGR）作为因子筛选顺序的依据，依次去除信息增益率较小的因子，直到基学习器多数投票结果准确率相对于基准较低为止。

信息增益 (IG) 是评价特征对于系统的相对影响程度, 计算在该特征下的信息熵与原始信息熵的差值, 即代表该特征带给整个系统的信息增益。利用信息增益指标有利于衡量特征对于整体系统的贡献程度, 但信息增益选择特征时容易偏向取值多的变量。而信息增益率 (IGR) 则克服了该问题, 信息增益率为信息增益 (IG) 除以分裂信息度量, 即考虑了该特征数据分裂的广度和均匀程度后的信息增益。由于在剔除冗余因子的过程中, 因子剔除顺序对于因子选择存在较大影响, 因此, 将信息增益率从小到大排序并依次剔除。

不平衡样本是财务舞弊识别模型首先需要解决的问题, 因为其对于有监督机器学习任务有较大影响 (Haixiang et al., 2017)。在传统算法中, 由于优化目标的设置会导致算法过多的关注多数类样本, 从而使少数类样本的分类准确度下降。已有文献一般有两种处理方法, 一是寻找匹配样本, 但是对于匹配样本选取而言, 匹配标准 (如资产规模相近、收入相近、净利润相近等)、匹配样本比例等参数确定都存在一定主观性, 并且不同参数组合对应的模型不同、预测效果也不同, 存在后验偏差且不符合真实预测场景。一般不平衡样本处理方式包括欠采样、过采样, 本文采用过采样的随机抽样算法对数据不平衡问题进行处理。

## (二) 元学习框架

对于训练集、验证集、测试集划分, 本文分别从正常样本和舞弊样本中随机抽取 1/3 样本组成测试集, 剩下的 2/3 样本作为调参数据集, 用 5 折交叉验证方式进行参数网格最优搜索, 用验证集表现最佳的参数作为最优模型。本文涉及的基学习器包括: 决策树、梯度提升树 (GBDT)、K 近邻算法、Logistic 回归、朴素贝叶斯算法、随机森林算法、支持向量机、极限梯度提升树 (XGBoost)、神经网络 (Neural Network), 每个基学习器遵循该流程进行调参。

元学习框架最好选取不同特征的基学习器 (Lemke et al., 2015), 因此, 我们选取了文献中常见的有监督分类算法作为基学习器。下文将对这几个算法进行简单说明。

第一类是线性分类算法。Logistic 回归是最常见的机器学习分类方法, 是一种线性回归分析模型, 适合二分类问题, 可以通过调整惩罚项控制过拟合程度。支持向量机 (SVM) 也是一种线性回归分析模型。与 Logistic 回归模型不同在于, 支持向量机侧重于最大化分类间隔, 而 Logistic 回归侧重于最小化分类误差, 即同一个问题下, 给定正则化系数, 如果用梯度下降法计算求 Logistic 回归最优, 参数初值会影响分类的超平面, 出现“一题多解”现象; 而 SVM 模型基于最大化分类间隔则不会出现多解情况。支持向量机还可以通过高斯核进行数据升维转换, 将线性不可分转换为非线性可分。



第二类是基于概率分布的算法。朴素贝叶斯算法假设每一个特征对于给定分类对其他特征对类别判定是相互独立的，但存在的问题是其分类结果受到初始假定数据  $Y$  分布的影响。因此，其数据分布假设是可调参数。

第三类是惰性算法。K 近邻算法 (kNN) 是最优代表性的惰性学习算法。kNN 原理在于用未知标签数据对应已知标签数据的距离，采取多数投票原则，其标签为多数邻居标签。K-近邻算法 (kNN) 优点是对异常值不敏感，但缺点是复杂度高，随着空间维度上升而上升，并且随着特征维数增加，样本在高维空间分布稀疏，导致“近邻”并不近，降低了预测效果。

第四类是以决策树为核心的算法，包括决策树、随机森林算法、梯度提升树 (GBDT)、极限梯度提升树 (XGBoost)，它们分别是决策树、决策树结合 bagging 算法、结合 boosting 梯度提升以及目标函数优化得到的。决策树模型是对应每一个叶子节点，利用特征值将数据集分为两个部分，若其中一个叶子节点不纯度在阈值以下，则不继续用特征值对该叶子节点继续划分。随机森林算法、梯度提升树 (GBDT)、极限梯度提升树 (XGBoost) 都属于基学习器为决策树的集成学习算法。一般限制的参数也较多，例如最大树深度、颗数限制、特征划分标准等。

第五类是神经网络 (Neural Network)。神经网络的核心是激活函数，通过多层神经元的链接构成网络以实现复杂非线性可分的问题。可调参数较多，例如隐藏层数、激活函数、优化权重器、学习率、最大迭代次数等。

以下部分介绍元学习框架的预测流程。设混合样本一共有  $k$  个，特征有  $m$  个，则元学习框架具体表示为：

- a) 设原始数据为  $k * m$  维向量  $X$ ，对应分类标签为  $k * 1$  维向量  $y$ ，将数据放入基学习器中训练，通过调参获得验证集预测准确度最高的模型，对训练集、验证集和测试集用最优模型预测得到  $z$ ；每一个分类器获得一个  $k * 1$  维的预测向量  $z$ ，将所有分类器预测结果堆叠形成  $k * 9$  维的  $Z$ ；
- b) 将  $Z$  从基学习器输出，输入到不同的堆叠分类器中再一次通过不同机器学习算法训练，对应分类标签依旧是  $k * 1$  维向量  $y$ ；
- c) 每一个堆叠分类器输出结果为  $k * 1$  维向量  $z'$ ，通过计算

$$y' = \begin{cases} 1, \text{avg}(Z') > 0.5 \\ 0, \text{avg}(Z') \leq 0.5 \end{cases} \quad (1)$$

获得预测值  $y'$ 。

通过元学习框架，整个流程能够自动学习各个分类器在给定情况下的优势和劣势，从底层模型的预测和分类偏差中学习，获得更大的分类能力 (Brazdil et al., 2009; Bhatt et al., 2013)。这种堆叠的成功来源于它可以利用基学习器的预测多样性，从而在元级获得更高的预测准确性，这种基学习器的再学习相较

于单个分类器或者简单结合策略会更有效（Abbasi et al., 2012），并且在支持数据挖掘自动化方面提高学习模型的泛化能力和稳定性（Parmezan et al., 2017）。

## 四、数据说明

本文模型中的输入数据从数据来源上划分，包括年报因子和季报因子；从类型上划分，包括财务类原始因子、财务类调整因子、公司治理因子、语言因子。

### （一）财务舞弊样本标记

本文采用国泰安数据库违规信息数据，其中包含了违规事件的证券代码、实际违规年份、违规类型的数据，其中违规类型包括虚构利润、虚列资产、推迟披露、出资违规、擅自改变资金用途、内幕交易、操纵股价、违规担保等。本文财务舞弊样本选取违规类型包括虚构利润、虚列资产、虚假记载（误导性陈述）、披露不实（其它）作为舞弊样本。

### （二）财务因子选取以及衍生因子构建

本文主要参照 Beneish（1999）、Cecchini 等（2010）、Abbasi 等（2012）以及文献中最常见的财务舞弊指标，从资产结构、盈利能力、现金流量、营运能力四个方面选取了 11 个公司基本面特征变量。资产结构方面，本文选取了资产质量指数、杠杆率（Beneish, 1999; Cecchini et al., 2010）。资产质量指数计算方式是剔除固定资产的非流动性资产/总资产，并将当期（t）除以上一期（t-1）。若资产质量指数>1 说明，该公司潜在的递延成本增加，有资产夸大的可能。杠杆率是当期总债务/总资产比上一期该值。若杠杆率上升，债务增长相对于资产增长过快，存在财务危机可能。

盈利能力方面，本文选取了边际利润率、净利润率、销售收入增长率（Beneish, 1999; Cecchini et al., 2010; Abbasi et al., 2012）。边际利润率是上一期边际利润率/当期边际利润率。当边际利润率>1 说明边际利润率在恶化，说明公司可能在虚增收入而利润没有变化。净利润率即净利润/营业利润，若企业虚增收入而没有对应的成本增加，会导致净利润上涨过快，净利润率过大。销售收入增长率是当期销售收入/上一期销售收入，若虚增销售收入，销售收入增长率会上涨过快。

现金流量方面，本文选取了净经营现金流与净利润的差值。净经营现金流与净利润的差值评估应计项目对财务报表的影响（Beneish, 1999）。该比率如果为正，说明存在潜在收入造假的可能。

营运能力方面，本文选取了总资产周转率、应收账款周转天数变化率、存货增长率、应收账款增长率、销售管理费用增长率（Cecchini et al., 2010; Abbasi et al., 2012）。总资产周转率是销售收入/总资产，当虚增销售收入，会导致该值偏大。应收账款周转天数变化率是当期（t）应收账款/销售收入比上一期（t-1）期数据，若企业虚增收入，会导致应收账款虚增，那么应收账款周转天数也会增加。当期存货/上一期存货。当存货增长率值越大，说明部分销售成本转嫁到存货账面成本，虚增营业利润。应收账款增长率是当期应收账款比上一期应收账款。虚增收入会导致应收账款过高，应收账款增长率增加过快。销售管理费用增长率是当期销售管理费用占销售收入比/上一期销售、管理费用占销售收入比。若公司虚增收入，那么销售、管理费用占比会下降，销售管理费用增长率会更小。

另外，考虑到行业偏差和结构偏差（Abbasi, 2012），本文在 11 个指标的基础上增加了衍生的行业调整因子和衍生的结构调整因子。首先是行业调整因子。行业调整因子分为两类，一个是样本对应同行业营业收入最接近的五家公司的对应因子平均，一个是同行业营业收入最高的前五样本因子平均。然后再将原有的公司特有因子分别与同行业营业收入最接近的五个公司平均、同行业营业收入前五平均对应因子做差（-）或者做除法（\）。其次是结构调整因子，即使用该因子对应的同比值，然后做差（-）或者做除法（\）。

### （三）治理因子构建

学术界研究发现上市公司高管背景、高管组织结构、股本结构等与会计信息质量、财务重述行为具有一定相关性（Lin et al., 2015; 卢馨等, 2015; Habib and Jiang, 2015）。本文选取了高管团队背景、财报审计、股本结构、治理结构几个方面构造公司治理因子。

高管背景方面，本文选取了高管团队女性性别占比、平均年龄、教育背景作为高管背景因子，数据来源国泰安数据库。Gao 等（2017）、Liao 等（2019）发现女性的存在与实施财务舞弊的低可能性相关，因此女性比例越高，财务舞弊可能性就越低。Hambrick 和 Mason 认为年龄大的高管人员更加厌恶风险，因此高管平均年龄有可能与财务舞弊行为呈负相关。卢馨等（2015）发现高管学历与财务舞弊行为严重负相关，说明学历越高财务舞弊可能性就越低。

财报审计方面，本文选取了是否所属四大审计事务所、当年年报审计意见、披露年报时间作为财报审计的特征因子。四大审计事务所基于丰富经验以及出于维持其声誉的考虑，其相较于其他审计事务所的审计报告质量更高且更可信。当年年报审计意见因子定义为标准无保留或者无保留意见加事项段为 1，保留或保留意见加事项段为 2，否定以及无法发表意见为 3。披露年报时间定义为，该公司财务报表正式披露

年报时间距离上一年 12 月 31 日之间间隔的天数。Conover 等（2008）发现财务报表披露时滞与资本市场监督程度存在相互关系，若财务报告披露越延迟，往往公司的业绩越欠佳，越有可能进行财务舞弊。

股本结构方面，本文选取了股本结构是否变化、董事会持股数量占比、管理层持股总股数、国有股股数占比、流通股数占比、监管层持股数、股东总数作为股本结构特征。许多学者都研究了股权结构对财务报表质量、公司治理能力的影响（Habib and Jiang, 2015; Gao et al., 2017）。总的来说，政府关联公司的治理水平较高，以及股权较分散、治理结构变动频率较低的公司财务舞弊可能性也相对较低。

治理质量方面，本文选取了前十大股东是否存在关联、董事长与总经理是否同一人、高管团队总人数、董事人数、独立董事人数占比、监事会人数占比、董事监事及高管年薪总额作为治理质量因子。已有研究发现董事会规模与财务舞弊有相关性，较高的独立董事占比意味着较高的盈余信息质量（胡奕明和唐松莲，2008），能够抑制财务舞弊事件的发生（Beasley, 1996; Dechow et al., 1996）。另外，高管年薪与财务舞弊也有一定关系，股票薪酬的正向激励与其财务舞弊的负面效应存在矛盾（Erickson et al., 2006; 方军雄，2012），而舞弊公司往往给予高管较低的现金薪资（Persons, 2006）。

#### （四）语言因子构建

已有学者发现，年报文字所蕴含或积极或消极的信息与公司业绩、高管行为、财务报表质量有显著关系。Goel 和 Uzuner（2016）、Petr 和 Roberto（2017）对美国上市公司董事会分析（MD&A）部分进行了文本分析，中国上市公司从 2005 年开始也包含了这一项内容，这一部分内容是基于管理层视角对于公司当前表现以及未来规划的文字叙述。已有文献发现，财务造假者倾向于采用负面以及不确定的词汇进行表述（Newman et al., 2003; Throckmorton et al., 2015），以及 MD&A 提及的对业绩产生负面影响的内部和外部因素越多，亏损扭转的可能性越小（薛爽等，2010）。并且已有证据表明，中国上市公司年报存在语调管理行为（曾庆生等，2018）。因此，本文将基于 MD&A 部分的文本进行语言分析，包括 7 个指标：正向词、负向词、情感基调、强烈语气、模糊语气、确定性程度、表达观点动词。

正向、负向词按照知网正负面评价情感词库、台湾大学简体中文情感词典、清华大学李军中文褒贬词典进行分辨，强烈语气词按照知网程度级别词库中，“极其、最、很、超”类别进行分辨，模糊语气按照知网程度级别词库中“稍”类进行分辨，表达观点动词借助知网词库中主张词语进行分类。以下是具体语言因子的构造说明。

首先，从东方财富网上爬取 A 股财务报表（包括年报和季报）的董事会讨论与分析部分，并对 MD&A 的部分用 jieba 分词进行处理，其次，对 MD&A 整段文字长度进行统计（不包含标点符号）。

正向词因子（POS）为正向词出现次数/该段文本字符数，负向词因子（NEG）为负向词出现频数总和/该段文本字符数，同理可得强烈语气（STRONG）、模糊语气（UNCERT）、表达观点动词因子（REGARD）。

情感基调因子（TONE）用正向词因子和负向词因子组合而成，表示该段文字相对偏正向还是负向。

$$TONE = \frac{POS-NEG}{POS+NEG} \quad (2)$$

确定性程度因子（CERTAIN）用强烈语气与模糊语气词因子进行合成，表示该段文字的确定性程度。

$$CERTAIN = \frac{STRONG-UNCERT}{STRONG+UNCERT} \quad (3)$$

由于部分公司会出现更新财务报表的情况，本文采取的语言因子采用更新后的报表 MD&A 进行文本分析。并且中国 A 股从 2005 年才开始要求对这一部分进行强制披露，爬取文本出现较多格式错误，以及提取董事会讨论与分析文本部分规则较为混乱，因此文本因子数据从 2006 年开始。

## （五）数据说明

本文财务数据用国泰安数据库财务报表数据进行计算，文本因子从东方财富网公司公告中爬取。先剔除了包含缺失值的样本，并对上下 1% 的异常值进行了处理，从 2006 年到 2018 年，最终有 1786 个舞弊样本，12770 个非舞弊样本。并用随机抽样算法对不平衡数据进行过采样处理，训练出模型后，再带入真实数据进行模型性能测试。每只股票在每年所采用因子以及个数如下表：

表 1 因子列表

类型	分项	总数
财务因子及其衍生因子	净经营现金流与净利润差值、存货增长率、应收账款周转天数、应收账款增长率、总资产周转率、杠杆率、净利润率、资产质量、边际利润率、销售收入增长率、销售管理费用增长率	各 28 个，共 308 个
治理因子	高管团队女性性别占比、平均年龄、教育背景、是否所属四大审计事务所、当年年报审计意见、披露年报时间、股本结构是否变化、董事会持股数量占比、管理层持股总股数、国有股股数占比、流通股数占比、监管层持股数、股东总数、前十大股东是否存在关联、董事长与总经理是否同一人、高管团队总人数、董事人数、独立董事人数占比、监事会人数占比、董事监事及高管年薪总额	各 1 个，共 20 个
语言因子	不确定因子、强烈语气因子、负向词因子、确定程度因子、正向词因子、表示观点动词因子、情感基调	各 4 个，共 28 个

## 五、实证结果

### （一）评价指标

由于本文为识别财务舞弊问题，相对于非舞弊样本中能够正确预测多少非舞弊样本而言，真实世界更关心预测结果为非舞弊样本中，有多少样本是真正非舞弊的公司。同理，对于预测结果为舞弊样本中，有多少样本是真正舞弊的公司。这样的指标称为召回率（recall），召回率更关心预测结果中的正确率。

与召回率相对应的是准确率（precision），这二者一般在同一条件下相对呈反向变化。为了兼顾这二者指标，F1-score 作为一个综合指标可以综合考察准确率与召回率的平衡情况。若召回率和准确率相差较大，F1-score 将会相对较小。

$$F1 = \frac{2*precision*recall}{precision+recall} \quad (4)$$

另一个模型性能指标是 AUC 数值，即为 ROC 曲线与坐标轴包围成的面积。ROC 曲线是按照财务舞弊识别模型预测概率进行降序排列，并累积计算模型的假正率与真正率，获得斜向上的 ROC 曲线。若 ROC 曲线围成的面积越大，AUC 数值则越高，也说明模型分类效果越好。

## （二）实证结果

### 1. 信息增益率（IGR）值排序统计

每个样本在年度横截面数据中，一共有 356 个因子，最终筛选剩下 87 个因子。以下对 IGR 值降序排序前 30、前 60、前 87 个因子进行统计，观察什么种类的因子对于财务舞弊识别更有效。

表 2 数据来源统计

属性	排名前 30	排名前 60	筛选后	总数
季度	5.56%	13.10%	22.22%	252
年度	15.38%	25.96%	29.81%	104
总计个数	30	60	87	356

以上统计数据为排序靠前的因子中，来源于年报或者季报因子占其总体因子的比例，百分比含义为该属性因子占该属性原有总体因子的比例。从上表来看，年度因子的有效性更高，说明年报数据相对于季报数据而言对于识别舞弊样本更加有效。并且从最终挑选的因子池来看，年度因子占比高于季度因子占比接近 30% 远远高于季度因子 22.22% 的占比。

表 3 数据类型统计

	前 30	前 60	筛选后	总数
行业调整财务因子	7.95%	18.18%	26.14%	176
结构调整财务因子	5.68%	15.91%	27.27%	88
原始财务因子	11.36%	15.91%	22.73%	44
治理因子	20.00%	25.00%	25.00%	20
语言因子	7.14%	7.14%	7.14%	28
总计个数	30	60	87	356

在最终因子池中被采用率最高的是结构调整财务因子，其占总财务结构因子的 27.27%，其次是行业调整财务因子占初始因子 26.14%，紧接着是治理因子 25%，最后是原始财务因子 22.73% 以及语言因子 7.14%。从 IGR 值来看，单个财务因子对于识别舞弊样本起到了主要作用，并且经过行业调整以及结构调整的财务因子对于整个分类系统也有较大的贡献。对于语言因子而言，在最终的因子池中只有两个因子入选，分别是第三季度确定程度因子和第一季度确定程度因子，并且分别排名第 2 和第 3，这说明在第三季度和第一季度董事会分析与讨论中，文本部分可能会释放表示确定性程度的信号，而这一信号对于识别公司舞弊将会有一定帮助。

## 2. 各机器学习器表现

首先，在不筛选因子的情况下，各机器学习器表现如表 4。

表 4 筛选前机器学习器表现

	非舞弊精准率	舞弊召回率	舞弊精准率	F1	AUC
Logistic 回归	0.87%	12.35%	99.83%	21.97%	50.35%
朴素贝叶斯	90.59%	21.23%	18.14%	19.57%	54.36%
支持向量机	99.02%	11.13%	6.50%	8.21%	50.82%
k 近邻	81.45%	33.96%	68.20%	45.34%	74.82%
决策树	58.62%	15.21%	53.08%	23.65%	55.85%
随机森林	99.22%	26.47%	2.02%	3.75%	50.62%
梯度提升树	98.68%	43.62%	7.28%	12.48%	52.98%
极限梯度提升树	98.87%	39.50%	5.26%	9.29%	52.07%
神经网络	93.81%	15.76%	8.29%	10.86%	51.05%

注 1：非舞弊精准率=预测为非舞弊样本中真实为非舞弊样本数/所有预测为非舞弊的样本数（下同）。

注 2：舞弊召回率=真实为舞弊样本中预测正确的个数/真实所为舞弊的样本数（下同）。

注 3：舞弊精准率=预测为舞弊样本中真实为舞弊样本数/所有预测为舞弊的样本数（下同）。

通过信息增益率进行排序并依次剔除信息增益率较小的变量，若剔除后机器学习器多数投票结果的 AUC 值低于剔除前的 AUC 值，则停止剔除。重复以上步骤获得了筛选后的因子库，并得到了各机器学习器表现结果见表 5。

表 5 机器学习器表现

	非舞弊精准率	舞弊召回率	舞弊精准率	F1	AUC
Logistic 回归	66.90%	17.05%	48.66%	25.25%	57.78%
朴素贝叶斯	77.17%	18.69%	37.51%	24.95%	57.34%
支持向量机	97.03%	21.61%	7.29%	10.90%	52.03%
k 近邻	78.96%	34.84%	80.46%	48.63%	79.71%
决策树	78.94%	17.84%	32.70%	23.09%	55.82%
随机森林	100.00%	100.00%	0.11%	0.22%	50.06%
梯度提升树	96.93%	35.74%	12.21%	18.20%	54.57%
极限梯度提升树	94.92%	34.25%	18.92%	24.38%	56.92%
神经网络	0.00%	12.27%	100.00%	21.86%	50.00%

对比表 4 和表 5，从筛选前后预测效果来看，一方面，除了决策树和神经网络的 AUC 值有小幅下降外，在因子筛选过后其余机器学习器算法的 AUC 值都有了明显的提高；从 F1 值来看，除了随机森林算法外，其他算法的 F1 值都有了明显提高。另一方面，无论是因子筛选前还是筛选后，各机器学习器的非舞弊精确率要明显平均高于舞弊精确率，并且因子筛选前该结论更为突出，这可能是由于一些信息增益较低的因子可能存在噪声信息，干扰了机器学习器的预测。



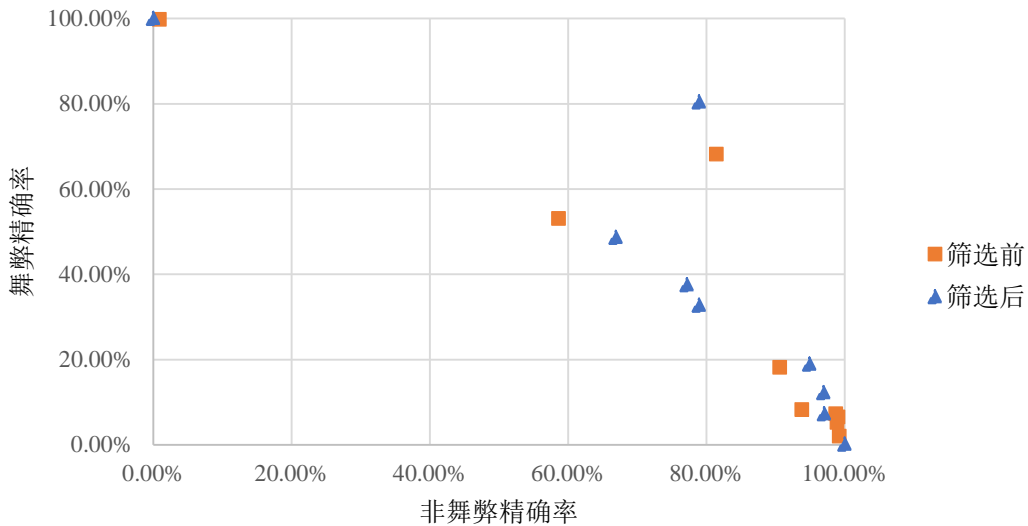


图 2 因子筛选前后机器学习器预测效果对比

另外，从机器学习器之间预测效果对比角度（图 2），无论在因子筛选前还是筛选后，Logistic 回归的舞弊精确率都要明显高于非舞弊精确率，这与其他机器学习器表现完全不同。除此之外，不同机器学习器的预测性能、非舞弊和舞弊精确率的权衡以及精确率和召回率的权衡上表现都不相同。对比筛选前后，不同算法在非舞弊和舞弊精确率的权衡分布也更为分散，这也为元学习框架学习吸收不同分类器的优势提供了有利条件。

### 3. 元学习框架下不同算法表现

将样本预测的结果进行堆叠，继续将预测结果作为解释变量，放入不同的学习器中进行进一步学习，得到如下结果：

表 6 元学习框架结果

	非舞弊精准率	舞弊召回率	舞弊精准率	F1	AUC
Logistic 回归	99.29%	95.09%	98.66%	96.84%	98.97%
朴素贝叶斯	98.65%	91.06%	98.71%	94.73%	98.68%
支持向量机	99.30%	95.14%	98.66%	96.87%	98.98%
k 近邻	99.22%	94.68%	98.66%	96.63%	98.94%
决策树	99.29%	95.09%	98.66%	96.84%	98.97%
随机森林	99.30%	95.14%	98.66%	96.87%	98.98%
梯度提升树	99.30%	95.14%	98.66%	96.87%	98.98%
极限梯度提升树	99.29%	95.09%	98.66%	96.84%	98.97%
神经网络	99.35%	95.47%	97.98%	96.71%	98.67%

由元学习框架结果对比各基学习器单层学习结果对比分析发现，无论是舞弊精准率、非舞弊精准率、舞弊召回率、AUC 值平均水平都有显著的提高。非舞弊精准率最高达到了 99.35%，最低精准率也达到了 98.65%，而舞弊精准率最低达到了 97.98%，最高则达到了 98.71%。另外，舞弊召回率也均分布在 91.06% 至 95.47% 的区间。从综合评价指标 F1 和 AUC 值来看，也明显高于基学习器之前的预测水平，F1 值从原来的 0.22%-48.63% 区间提高至将近 95% 以上的区间，AUC 指标从 50.00%-79.71% 区间提升值 98% 以上的区间，说明元学习框架对于提高预测舞弊样本具有显著的效果。

#### 4. 分行业观察元学习框架性能

在预测过程中，本文并未针对不同行业构建不同模型，即在元学习框架下并未对行业进行区分，对于所有行业均采用同一优化参数，那么元学习框架是否对于不同行业也有同样的适用度呢？

下表展示基于同一模型下，不同行业预测准确度是否表现良好。一些行业由于其其在 A 股上市的规模较大，其舞弊对中国资本市场造成的影响也会更大，因此，本文将按照证监会行业分类，重点观察上市企业较多的行业，分析其对应的舞弊识别性能如何。

表 7 分行业模型预测效果

	非舞弊精准率	舞弊召回率	舞弊精准率	F1	AUC	行业样本数
制造业	99.26%	95.01%	98.87%	96.90%	99.07%	9312
信息传输、软件和信息技术服务业	99.22%	95.35%	98.40%	96.85%	98.81%	959
批发和零售业	99.18%	93.64%	99.04%	96.26%	99.11%	894
房地产	99.54%	96.25%	97.47%	96.86%	98.50%	727
电力、热力、燃气及水的生产和供应业	99.32%	95.95%	98.61%	97.26%	98.97%	513
交通运输、仓储和邮政业	100.00%	100.00%	100.00%	100.00%	100.00%	381
建筑业	99.18%	93.94%	96.88%	95.38%	98.03%	324
采矿	99.29%	95.24%	95.24%	95.24%	97.26%	277
租赁和商务服务业	97.28%	90.91%	95.24%	93.02%	96.26%	204

本文选取了有效样本数排名前 10 的行业，可以看到 F1 分数均在 94% 以上，AUC 均 98% 以上。制造业作为周期性行业，业绩波动幅度较大，有较强的盈余管理以及财务造假动机，但制造业在元学习框架下，预测结果为舞弊样本中有 98.87% 为真正的舞弊样本，并且 F1 分数也达到了 96.90%，AUC 数值达 99.07%，说明元学习框架能够覆盖和有效预测制造业样本。对于有效样本前十的行业中，舞弊样本召回率最小值也高达 90.91%，说明大部分样本中，真实为舞弊样本的公司中有高达 90% 以上的样本预测正确。

不同行业财务舞弊的手段可能是不相似的，但是粉饰业绩的指标可能是相似的，例如应收账款周转率、销售收入等。虽然对于所有行业仅采用了一套优化参数，但是对于大部分行业的预测效果都较为稳定且良好，也说明了元学习框架在行业层面具有一定的稳健性。

### 5. 滚动预测性能分析

前面的结果按照全样本进行随机抽样组成训练集、验证集和测试集，但真实世界下，当年只能知道去年全年的信息，无法预知未来的舞弊信息以及财务信息，因此，该部分用两种滚动的方式检验元学习框架的稳健性。

第一种是以五年为单位进行滚动，例如 2006-2010 年为一组，2011 年作为测试集，2006 年-2000 年用 5 折交叉验证进行参数调参，然后每年都以固定时间 5 年为窗口进行滚动。第二种是时间窗口不断扩大的方式滚动，例如一开始以 2006-2010 年为一组，最新的 2011 年样本为测试集，2006 年-2010 年用 5 折交叉验证的方式进行参数调参；接着以 2006-2011 年为一组，最新的 2012 年测试集，2006 年-2011 年用 5 折交叉验证的方式进行参数调参，以此类推。

表 8 固定时间 5 年窗口滚动基学习器多数投票结果

	非舞弊精准率	舞弊召回率	舞弊精准率	F1	AUC
2010	68.06%	73.21%	84.95%	78.65%	69.41%
2011	63.49%	73.27%	89.94%	80.76%	69.52%
2012	59.15%	70.21%	88.54%	78.32%	68.08%
2013	59.79%	70.21%	86.97%	77.70%	67.56%
2014	65.46%	70.13%	80.73%	75.06%	67.06%
2015	65.55%	69.86%	80.63%	74.86%	67.56%
2016	67.97%	69.80%	77.54%	73.47%	67.17%
2017	69.69%	70.53%	76.90%	73.58%	67.39%
2018	70.21%	71.34%	77.63%	74.35%	68.10%

注：滚动结果是样本外测试集预测结果（下同）。

表 9 固定时间 5 年窗口滚动元学习多数投票结果

	非舞弊精准率	舞弊召回率	舞弊精准率	F1	AUC
2010	97.14%	97.19%	98.53%	97.85%	86.98%
2011	91.11%	92.63%	99.06%	95.73%	87.03%
2012	88.32%	89.99%	98.58%	94.09%	86.45%
2013	88.08%	90.31%	98.25%	94.11%	86.55%
2014	90.76%	91.29%	95.83%	93.50%	86.24%
2015	93.54%	93.74%	96.44%	95.07%	86.35%
2016	89.85%	90.59%	97.32%	93.83%	85.89%
2017	93.04%	93.35%	97.52%	95.39%	86.26%
2018	93.03%	93.47%	98.50%	95.92%	86.65%

对比表 8 和表 9，元学习框架能够较好的提高舞弊样本召回率，舞弊样本准确率也有一定的提升，说明元学习框架对于基学习器预测效果提升有较大的帮助。虽然对比随机抽样生成的训练集、验证集、验证集而言，按照真实世界采用固定窗口滚动预测方法的 AUC 值有所下降，但最低也达到了 86.24%，F1 最低也达到了 93.50%，并且相较于单层基学习器的多数投票预测结果而言，明显有显著的提高。

我们尝试第二种时间窗口不断扩大滚动的预测方式也得到了相同的结论（表 10、表 11）。

表 10 时间窗口不断扩大滚动基学习器多数投票结果

	非舞弊精准率	舞弊召回率	舞弊精准率	F1	AUC
2010	68.06%	73.21%	84.95%	78.65%	69.41%
2011	69.07%	73.37%	83.87%	78.27%	69.02%
2012	62.22%	71.67%	87.30%	78.71%	68.50%
2013	67.71%	71.07%	80.65%	75.56%	67.56%
2014	66.90%	70.88%	81.01%	75.61%	66.33%
2015	57.97%	69.40%	87.39%	77.36%	67.13%
2016	60.98%	68.95%	83.29%	75.45%	66.77%
2017	57.25%	68.76%	87.03%	76.82%	66.82%
2018	68.04%	70.74%	78.74%	74.53%	67.36%

表 11 时间窗口不断扩大滚动元学习多数投票结果

	非舞弊精准率	舞弊召回率	舞弊精准率	F1	AUC
2010	97.14%	97.19%	98.53%	97.85%	86.98%
2011	98.29%	98.31%	99.53%	98.91%	87.22%
2012	92.30%	93.23%	97.91%	95.51%	86.79%
2013	91.10%	91.79%	97.66%	94.63%	86.47%
2014	95.36%	95.52%	98.73%	97.10%	86.75%
2015	94.10%	94.33%	97.56%	95.91%	86.57%
2016	88.49%	89.95%	97.41%	93.53%	86.34%
2017	91.47%	92.09%	98.14%	95.02%	86.63%
2018	95.34%	95.47%	97.82%	96.63%	86.79%

对比固定时间窗口（表 8、表 9）的滚动以及扩大时间窗口（表 10、表 11）的滚动结果，基学习器表现以及元学习框架表现大体相差较小，说明舞弊企业特征与非舞弊企业特征的识别模式没有较大的改变。并且在元学习框架下，不同年份之间识别的精准率、召回率、F1 值都维持在 90% 以上，AUC 值维持在 85% 以上，并未有太大的波动，也侧面说明了元学习框架预测效果的稳健性和预测性能的优越性。

## 六、结论与展望

本文对于无缺失数据的 2437 家上市公司进行分析，涉及三大类五小类包括原始财务因子、行业调整财务因子、结构调整财务因子、公司治理因子、语言因子共计 356 个因子，并依次对因子进行筛选、异常值处理，再对数据不平衡问题处理，最后运用元学习框架对基学习器进行优化改进，以提高上市公司财务舞弊识别的准确性，并得到了如下结论：

第一，元学习框架能够有效的提升财务舞弊识别的准确性，可以体现在提升非舞弊样本和舞弊样本精确率、舞弊样本召回率、F1、AUC 数值。并且元学习框架对于不同行业也同样适用，有效样本排名前十的行业召回率均在 90% 以上，说明了元学习框架的广泛适用性。

第二，模拟真实世界的信息流，采用滚动预测的方式对元学习框架进行测试，虽然结果稍逊于随机抽样生成测试集的结果，但召回率均维持在 90% 以上，AUC 数值均维持在 85% 以上。并且对比不同滚动方式的预测结果，基学习器表现以及元学习框架表现大体相差较小，说明舞弊企业特征与非舞弊企业特征的识别模式没有较大的改变。

第三，年度财务报告中提取的因子相对于季度报告因子，在识别舞弊方面更有效。财务类因子中，结构偏差调整因子对于财务舞弊识别帮助最大，其次是行业偏差调整因子，最后是原始因子。其次，最终有

25%的公司治理因子被纳入最终因子库，具体包括董事会持股比例、是否为四大审计事务所、审计意见、报告披露时间、高管薪资。语言类因子对于财务舞弊识别也有一定的帮助，其中确定性程度因子对于财务舞弊识别帮助最大。

未来工作可以从以下两个方面进行深入研究。一方面，机器学习器可以进一步扩展，例如使用深度学习以进一步提高模型的预测效果。另一方面，本文仅将管理层经营讨论与经营部分进行了文本词性分析，今后可以对年报其他内容进行进一步分析并丰富文本分析方法。

## 参考文献

- (1) Abbasi, A., Albrecht, C., Vance, A., and Hansen, J., 2012, “Metafraud: a meta-learning framework for detecting financial fraud” , *MIS Quarterly*, Vol.36(4), pp.1293~1327.
- (2) Albashrawi, M., and Lowell, M., 2016, “Detecting financial fraud using data mining techniques: a decade review from 2004 to 2015” , *Journal of Data Science*, Vol.14(3), pp.553~569.
- (3) Beasley M., 1996, “An empirical analysis of the relation between board of director composition and financial statement fraud” , *Accounting Review* , Vol.71, pp.443~466.
- (4) Beasley, M. S., Carcello, J. V., Hermanson, D. R. and Neal, T. L., 2010, “Fraudulent financial reporting: 1998-2007: An analysis of US public companies”, *Committee of Sponsoring Organizations of the Treadway Commission*.
- (5) Bell, T. B., and Carcello, J. V., 2000, “A decision aid for assessing the likelihood of fraudulent financial reporting” , *Auditing: A Journal of Practice and Theory*, Vol.19(1), pp.169~184.
- (6) Beneish, M. D., 1999, “The detection of earnings manipulation” , *Financial Analysts Journal*, Vol.55(5), pp.24~36.
- (7) Bhatt, N., Thakkar, A., Ganatra, A., Bhatt, N., 2013, “Ranking of classifiers based on dataset characteristics using active meta learning” , *International Journal of Computer Applications*, Vol.69 (20), pp.31~36.
- (8) Bologna, J., Lindquist, R. J., and Wells, J. T., 1993, “The accountant's handbook of fraud and commercial crime, *New York, NY:Wiley*.
- (9) Brazdil, P., Carrier, C. G. G., Soares, C., and Vilalta, R., 2009, “Metalearning - applications to data mining” , *Cognitive Technologies*.
- (10) Cecchini, M., Aytug, H., Koehler, G. J., Pathak, P., 2010, “Making words work: Using financial text as a predictor of financial events” , *Decision Support Systems*, 2010, Vol.50(1), pp.164~175.
- (11) Chen, G., Firth, M., Gao, D. N., and Rui, O. M., 2006, “Ownership structure, corporate governance and fraud: Evidence from China” , *Journal of Corporate Finance*, Vol.12(3), pp.424~448.

- (12) Conover, C. M., Miller, R. E., and Szakmary, A., 2008, “The timeliness of accounting disclosures in international security markets” , *International Review of Financial Analysis*, Vol.17(5), pp.0~869.
- (13) Cui, C., Hu, M., Weir, J. D., and Wu, T., 2016, “A recommendation system for meta-modeling: A meta-learning based approach” , *Expert Systems with Applications*, Vol.46, pp.33~44.
- (14) Dechow, P. M., Ge, W., Larson, C. R., and Sloan, R. G., 2011, “Predicting material accounting misstatements” , *Contemporary accounting research*, Vol.28(1), pp.17~82.
- (15) Dong, W., Liao, S., and Liang, L., 2016, “Financial Statement Fraud Detection using Text Mining: a Systemic Functional Linguistics Theory Perspective” , *PACIS 2016 Proceedings*, pp.188.
- (16) Dyck, A., Morse, A., and Zingales, L., 2010, “Who blows the whistle on corporate fraud?” , *The Journal of Finance*, Vol.65(6), pp.2213~2253.
- (17) Erickson, M., Hanlon, M., and Maydew, E. L., 2006, “Is there a link between executive equity incentives and accounting fraud?” , *Journal of Accounting Research*, Vol.44(1), pp.113~143.
- (18) Gao, Y., Kim, J. B., Tsang, D., and Wu, H., 2017, “Go before the whistle blows: an empirical analysis of director turnover and financial fraud” , *Review of Accounting Studies*, Vol.22(1), pp.320~360.
- (19) Glancy, FH., Yadav, SB., 2011, “A computational model for financial reporting fraud detection”, *Decision Support Systems*, Vol.50, pp.595~601.
- (20) Goel, S., and Uzuner, O., 2016, “Do sentiments matter in fraud detection? Estimating semantic orientation of annual reports” , *Intelligent Systems in Accounting, Finance and Management*, Vol.23(3), pp.215~239.
- (21) Habib, A., and Jiang, H., 2015, “Corporate governance and financial reporting quality in China: A survey of recent evidence” , *Journal of International Accounting, Auditing and Taxation*, Vol.24, pp.29~45.
- (22) Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., and Bing, G., 2017, “Learning from class-imbalanced data: Review of methods and applications” , *Expert Systems with Applications*, Vol.73, pp.220~239.
- (23) Hajek, P., and Henriques, R., 2017, “Mining corporate annual reports for intelligent detection of financial statement fraud—A comparative study of machine learning methods” , *Knowledge-Based Systems*, Vol.128, pp.139~152.
- (24) Hambrick D C, Mason P A., 1984, “Upper echelons: the organization as a reflection of its top managers” , *Academy of Management Review*, Vol.9(2), pp.193~206.
- (25) Hasnan, S., Rahman, R. A., and Mahenthiran, S., 2012, “Management motive, weak governance, earnings management, and fraudulent financial reporting: Malaysian evidence” , *Journal of International Accounting Research*, Vol.12(1), pp.1~27.

- (26) Hastie, T., Tibshirani, R., Friedman, J., and Franklin, J., 2005, “The elements of statistical learning: data mining, inference and prediction” , *The Mathematical Intelligencer*, Vol.27(2), pp.83~85.
- (27) Hoogs, B., Kiehl, T., Lacombe, C., and Senturk, D., 2007, “A genetic algorithm approach to detecting temporal patterns indicative of financial statement fraud”, *Intelligent Systems in Accounting, Finance and Management* , Vol.15, pp.41~56.
- (28) Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., and Felix, W. F., 2011, “Identification of fraudulent financial statements using linguistic credibility analysis” , *Decision Support Systems*, Vol.50(3), pp.585~594.
- (29) Lemke, C., Budka, M., and Gabrys, B., 2015, “Metalearning: a survey of trends and technologies” , *Artificial intelligence review*, Vol.44(1), pp.117~130.
- (30) Liao, J., Smith, D., and Liu, X., 2019, “Female CFOs and accounting fraud: Evidence from China” , *Pacific-Basin Finance Journal*, Vol.53, pp.449~463.
- (31) Lin, C. C., Chiu, A. A., Huang, S. Y., and Yen, D. C., 2015, “Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts’ judgments” , *Knowledge-Based Systems*, Vol.89, pp.459~470.
- (32) Matijaš, M., Suykens, J., and Krajcar, S., 2013, “Load forecasting using a multivariate meta-learning system” , *Expert Systems with Applications*.
- (33) Newman, M. L., Pennebaker, J. W., Berry, D. S., and Richards, J. M., 2003, “Lying words: Predicting deception from linguistic styles”, *Personality and social psychology bulletin*, Vol.29(5), pp.665~675.
- (34) Parmezan, A. R. S., Lee, H. D., and Wu, F. C., 2017, “Metalearning for choosing feature selection algorithms in data mining: Proposal of a new framework” , *Expert Systems with Applications*, Vol.75, pp.1-24.
- (35) Persons, O. S., 2006, “The effects of fraud and lawsuit revelation on US executive turnover and compensation” , *Journal of Business Ethics*, Vol.64(4), pp.405~419.
- (36) Pincus, K. V., 1989, “The efficacy of a red flags questionnaire for assessing the possibility of fraud. Accounting” , *Organizations and Society*, Vol.14(1-2), pp.153~163.
- (37) Purda, L., and Skillicorn, D., 2015, “Accounting variables, deception, and a bag of words: Assessing the tools of fraud detection” , *Contemporary Accounting Research*, Vol.32(3), pp.1193~1223.
- (38) Ravisankar, P., Ravi, V., Rao, G. R., and Bose, I., 2011, “Detection of financial statement fraud and feature selection using data mining techniques” , *Decision Support Systems*, Vol.50(2), pp.491~500.
- (39) Seo, J. H., and Choi, D., 2016, “Feature selection for chargeback fraud detection based on machine learning algorithms” , *International Journal of Applied Engineering Research*, Vol.11(22), pp.10960~10966.



- (40) Song, X. P., Hu, Z. H., Du, J. G., and Sheng, Z. H., 2014, "Application of machine learning methods to risk assessment of financial statement fraud: evidence from China", *Journal of Forecasting*, Vol.33(8), pp.611~626.
- (41) Throckmorton, C. S., Mayew, W. J., Venkatachalam, M., and Collins, L. M., 2015, "Financial fraud detection using vocal, linguistic and financial cues", *Decision. Support Systems*, Vol.74, pp.78~87.
- (42) Vilalta, R., and Drissi, Y., 2002, "A perspective view and survey of meta-learning", *Artificial Intelligence Review*, Vol.18(2), pp.77~95.
- (43) West, J., and Bhattacharya, M., 2016, "Intelligent financial fraud detection: a comprehensive review", *Computers and security*, Vol.57, pp.47~66.
- (44) Wilks, T. J., and Zimbelman, M. F., 2004, "Decomposition of fraud-risk assessments and auditors' sensitivity to fraud cues", *Contemporary Accounting Research*, Vol.21(3), pp.719~745.
- (45) 曾庆生、周波、张程、陈信元：《年报语调与内部人交易：“表里如一”还是“口是心非”？》，《管理世界》，2018年第9期。
- (46) 方军雄：《高管超额薪酬与公司治理决策》，《管理世界》，2012年第11期。
- (47) 胡奕明、唐松莲：《独立董事与上市公司盈余信息质量》，《管理世界》，2008年第9期。
- (48) 李丹、宋衍蘅：《及时披露的年报信息可靠吗？》，《管理世界》，2010年第9期。
- (49) 薛爽、肖泽忠、潘妙丽：《管理层讨论与分析是否提供了有用信息？——基于亏损上市公司的实证探索》，《管理世界》，2010年第5期。
- (50) 朱锦余、高善生：《上市公司舞弊性财务报告及其防范与监管——基于中国证券监督管理委员会处罚公告的分析》，《会计研究》，2007年第11期。