

# 基于模糊神经网络的金融仓单风控数据筛选系统

顾凌云

(上海冰鉴信息科技有限公司, 上海 200120)

**摘要:** 现有的数据筛选系统使用方法实时性较低, 影响结果置信度, 无法满足特定场景下金融仓单风控的需要。因此设计一种基于模糊神经网络的风控数据筛选系统。将仓单页面表示为二维矩阵, 经过离散余弦变换将仓储量和仓单类型进行转换, 实现金融仓单页面精准分析, 对金融仓单中数据进行特征提取, 利用数据集中分类的属性方程对风控数据进行特征向量匹配, 分析模糊神经网络的输入和输出数据, 对所有的数据进行遍历, 得到风控数据结果。在系统性能测试中, 设计的系统输出的数据具有很强的特征聚类性。

**关键词:** 模糊神经网络; 金融仓单; 数据筛选; 特征匹配

中图分类号: TP183; TP391.3 文献标识码: A 文章编号: 1003-7241(2025)03-0074-05

## Financial Warehouse Receipt Risk Control Data Screening System Based on Fuzzy Neural Network

GU Ling-yun

(Shanghai Icecredit, Inc., Shanghai 200120 China)

**Abstract:** The existing data filtering system has low real-time performance, which affects the confidence of the results and cannot meet the needs of risk control of financial warehouse receipts in specific scenarios. Therefore, this paper designs a risk control data screening system based on fuzzy neural network. The warehouse receipt page is represented as a two-dimensional matrix, and the warehouse reserves and warehouse receipt types are converted through discrete cosine transformation to achieve accurate analysis of the financial warehouse receipt page, feature extraction of the data in the financial warehouse receipt, feature vector matching of the risk control data using the attribute equation classified in the data set, analysis of the input and output data of the fuzzy neural network, and traversal of all data to obtain the risk control data results. In the system performance test, the output data of the designed system has strong feature clustering.

**Keywords:** fuzzy neural network; financial warehouse receipt; data filtering; feature matching

### 0 引言

金融仓单作为一种新型的仓储交易和抵押方法, 随着互联网应用的普及, 受到各银行和仓储企业的广泛应用。但银行为了规避相应的风险, 一般会选择固定资产、重金属货物等作为融资对象<sup>[1-2]</sup>。而中小微企业这种类型的抵押产品较小, 通常为大宗类产品, 产品种类较多, 产品价格与当前市场价格有紧密联系。银行受限于技术限制, 难以对所有货物的市场价格进行统计, 也无法对抵押货物进行合理的估值, 造成潜在的金融交易风险<sup>[3-4]</sup>。解决大宗类商品的货物估值问题, 首先需要获取市场上该商品的价格信息, 但由于海量数据、信息准确提取等因素的限制, 目前还没有针对金融仓单类型的高效网络爬虫算法<sup>[5-7]</sup>。目前国内银行和期货公司等金融机构所使用的数据, 主要来源于仓库实际管理人员和工作人员, 耗费率

力物力的同时, 效率和准确性都存在巨大的限制。另一方面, 由于金融仓单中货物价格具有一定的时效性, 在某些特殊情况下, 如易变质货物的估值等, 对实时性要求非常高, 因此必须采用并行或分布式爬虫技术。目前, 国内外网络爬虫技术的算法中, 代表算法有鱼群搜索策略、鲨鱼搜索策略与最佳优先搜索策略。以上几种代表性算法具有搜索速度受限、受相似文本影响、遗漏等问题。同时, 现有网络爬虫技术中必然用到关键字、已爬取 URL 列表等资源的共享和更新操作, 采用并行的方法必然存在共享资源的单节点的热点资源不足问题, 这些问题导致现在有的并行(分布式)爬虫方法, 并不能使性能产生线性增长, 导致爬虫的实时性较低, 造成网络传输的时延, 影响数据筛选的结果置信度, 无法满足特定场景下金融仓单风控的需要。因此本文采用了神经网络的分类模型, 并结合模糊神经网络的算法, 设计一套基于模糊神经网络的优化风控数据筛选流程。

\*基金项目: 上海市软件和集成电路产业发展专项资金项目 (200316)

收稿日期: 2023-03-31

## 1 金融仓单风控数据筛选系统设计

### 1.1 金融仓单页面分析

金融仓单是一种有价证券,由上海金属交易所、郑州商品交易所、大连商品交易所等机构发行,其代表指定的库存物品的所有权。持有人可以通过交易金融仓单来获得该物品的所有权,并享有与该物品相关的收益和利益。这种证券的出现,使得一些银行、贸易公司、投资机构等可以通过金融仓单进行融资和投资。金融仓单风控数据通常包括以下方面:

(1) 金融仓单数量:即某一品种的仓单数量,这个数据可以反映该品种的需求与供给情况;

(2) 金融仓单库存:即某一品种的金融仓库库存,用于评估该品种市场供应与需求;

(3) 金融仓单品质:有的金融仓单品种有不同的规格和品质,针对不同品质的金融仓单,需要有区别对待的风险控制措施;

(4) 金融仓单的历史走势:跨时间段展示某一品种或交易所的金融仓单总量和变化情况,用于了解该品种市场历史走势,为风险控制提供参考。

这些金融仓单风控数据是反映金融市场上某一品种的供求关系、市场趋势等的重要指标,供金融机构、投资人等基于风险控制需求分析市场风险,及采取相应的投资策略与决策。

对于一般的金融仓单来说,主要有以下类型:商品现货仓单、租赁仓单。分析仓储公司的类型,确定仓储公司的性质是否为银行或期货公司的分支机构,以及仓库的分布是否是保税仓<sup>[7-8]</sup>。在实际的金融仓单页面的分析过程中,需要考虑数据源的选择依据,数据源选择是一个很重要的问题,要根据自身的业务情况来考虑具体的应用场景。一个仓单页面在数字中可以表示为二维矩阵,该二维矩阵图像的离散余弦变换公式如下<sup>[8]</sup>:

$$F(u,v) = \alpha(u)\alpha(v) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \cos \frac{(2x+2y)u\pi}{2M \times N} \quad (1)$$

式中, $f(x,y)$ 表示图像中某点坐标为 $(x,y)$ 的像素值, $M \times N$ 为二维矩阵图像尺寸,式中的 $\alpha(u)$ 和 $\alpha(v)$ 可以转换为:

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{M}} & u = 0 \\ \sqrt{\frac{2}{M}} & u \neq 0 \end{cases} \quad (2)$$

$$\alpha(v) = \begin{cases} \sqrt{\frac{1}{N}} & v = 0 \\ \sqrt{\frac{2}{N}} & v \neq 0 \end{cases} \quad (3)$$

金融机构需要从多个维度来进行考察,如业务背景、仓储量、交易对手、仓库所属区域等,这是从数据来源角

度来考虑的;对于仓储公司,需要考察其历史业绩、经营管理能力等情况,这是对仓储公司发展的一种动态评估<sup>[9]</sup>。评估过程中,需要进行哈希值计算,URL根据不同的哈希值发送到不同区段节点。对于仓储量需要考察其是否存在多家分公司,是否具备多家合作仓库以及其他类似情况可以提供给模型使用;对于交易对手,需要考察在近几年是否有类似情况发生(如连续几个月未交易),是否具备相关能力;仓库所属区域需要分析其所在城市以及仓库分布的详细情况。

在获取到以上数据之后,对得到的数据进行预处理。对于数据预处理的要求主要体现在两个方面:一是数据的完整性,即原始数据应该包括各个要素,例如仓储量、仓单类型、仓储费等;二是数据的及时性,即从原始数据库中抽取的仓储量和仓单时要与业务过程中产生的原始数据库保持一致,以保证数据的及时性。由于期货业务中存在许多仓储量和仓单之间有明显差异情况,为了保证分析时在训练阶段进行快速学习及验证结果准确可靠,在该模型中将所有仓储量和仓单类型之间采用以下方法进行转换:

(1) 将其中一个指标设置为变量值为0,其他指标设置为变量值=1;

(2) 对于所有仓容量和仓单类型之间可以直接进行转换;

(3) 通过对已有仓储费的使用分析,获取仓储费训练效果最好时的占比,可以通过改变仓储量(增加或减少)、增加仓单类型、降低仓储量、改变仓储量的方式来实现模型的快速学习和验证,实现金融仓单页面的精准分析。

### 1.2 风控数据特征向量匹配

由于仓单所包含的信息极为丰富,可以通过对这些信息进行筛选,得到更多的价值信息<sup>[10-12]</sup>。在实际数据集的构建过程中,首先对所选取的每一个样本集都做一个标签化处理,即数据特征提取,以便于在模型训练和测试阶段,将每一张数据集中所对应的标签信息作为输入变量,并对该样本集进行模糊编码,进而将不同类型的指标转化为不同的权重向量。确定仓单属性,针对每个仓库,首先要对其入库的所有仓单进行检索,根据每个仓库的历史表现(包括入库时间、库龄、货物价值等)和特征,确定每个仓单的基本信息以及具体特征值(如:名称、代码、编码;是否有特殊标识;是否是新入库不久等)。提取不同仓库中最具代表性的数据进行提取。主要是指以下几个方面:仓库库存明细、历史数据等相关数据。在获取到以上数据之后,根据这些基本信息建立初步模型。对原始数据进行预处理,提取其中的本体信息和特征匹配规则后,得到数据集中分类的属性方程:

$$F_z(t_k^0) = F(T_1 + T_2 + \dots + T_n \leq t_k^0) \quad (4)$$

式中,  $Z$ 表示系统中风控数据自动筛选节点的关联度, 构建替代函数, 得到特征中第  $j$  个特征属性的数值公式如下:

$$f(x) = SINR - 20 \exp\left(-0.2 \sqrt{\frac{F_z}{M \times N} \sum_{d=1}^{\alpha} x_d^2}\right) \quad (5)$$

式中,  $SINR$  是簇中心分布阈值,  $\alpha$  是不小于 1 的数, 将预处理后的特征向量带入到模糊神经网络中进行训练, 获取数据集的规模和簇中心的数量, 得到最终模型所需要的输入向量, 此时需要对类别特征向量进行匹配计算, 在类别特征向量库中筛选超过预制范围的特征, 将其锁定并更新, 直至所有特征都在阈值范围内, 此时将其他类别特征进行向量匹配计算, 再次利用阈值进行判断, 完成特征向量之间的匹配。

### 1.3 基于模糊神经网络优化风控数据筛选算法

在经过特征向量匹配之后, 对获得的每个特征向量结果采用相同参数进行模糊神经网络训练。假设输入的各个分量中, 模糊分割数据已经确定完成, 在进行模型训练和测试时, 首先利用上述标签化后的指标和数据作为输入变量建立模型。然后利用神经网络算法对输入变量及其隶属度计算公式进行学习训练和测试。那么在学习过程中需要确定的参数为连接权值和隶属函数内部系数。本文所使用的模糊神经网络属于多层前馈网络, 当输入数据为  $x_i$  时, 那么在模糊神经网络中的输出可以表示为:

$$f_i = \sum_{j=1}^m w_{ij} \omega^j f(x) \quad (6)$$

式中,  $\omega^T$  为标准差,  $w_{ij}$  为训练得到的模糊神经网络模型结果输出, 将训练结果按照一定规则进行处理得到最终模型结果。

将上文中经过匹配计算后的所有特征向量带入到模糊算法中, 在第三方软件中进行运算得到最终参数。计算出参数后将之带入到神经网络模型中并按照事先设定好的规则执行计算。对获得的所有特征向量进行标准化处理。

定义输入和输出的标准化公式:

$$\begin{cases} x = [x_1, x_2, \dots, x_n]^T \times \omega^T \\ y = [y_1, y_2, \dots, y_n]^T \times \omega^T \end{cases} \quad (7)$$

式中,  $x_i$  为原始样本数值,  $x$  为原始样本数值对应特征向量之和。  $y_i$  是原始样本数值对应特征向量,  $T$  是标准化公式中各个数据点到原始样本数值间隔时间后对应特征向量个数。

利用第三方软件对处理后的所有特征向量进行运算得到最终结果, 计算出每一个仓对应于某一货物所具有的特征值并将其放入相应类别中作为模型训练好之后所需要输出给出预测结果。总体的筛选流程如图 1 所示。

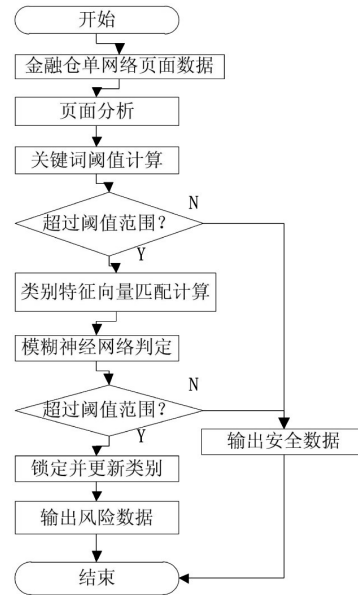


图 1 筛选流程示意图

如图 1 所示, 计算每个仓对相应属性和目标属性进行赋值, 并最终生成一个仓对应于某个特定目标属性的标签。通过以上步骤得到最终模型结果输出。至此完成金融仓单风控数据筛选系统的设计。

## 2 系统性能测试

### 2.1 实验参数设计

为了验证本文所设计的基于模糊神经网络的金融仓单封控数据筛选系统在实际应用中的有效性, 需要进行性能测试。实验主要采用 Matlab7 软件进行设计, 在该软件中, 分别搭载本文所设计的数据筛选系统、基于代码爬虫的数据筛选系统和基于粗糙集的数据筛选系统。测试过程中, 首先选择适合各个筛选系统的参数, 参数设置情况如表 1 所示。

表 1 系统参数设置

参数	本文数据筛选系统	基于代码爬虫的数据筛选系统	基于粗糙集的数据筛选系统
近义词簇关联度	50	50	50
嵌入维度	3	5	4
统计特征系数	5	3	3
邻域内拟合特征分布系数	6	5	4
采样幅值	1	1	1
信息熵	3	2	3
模糊度	0.03	/	/
能量函数	50	50	50
迁移学习目标阈系数	-70	20	30
语义结构相似性特征量	-30	10	10
系统局部信息熵	5	5	5

在测试过程中, 设置原始的分布式系统数据分布情况如图 2 所示。

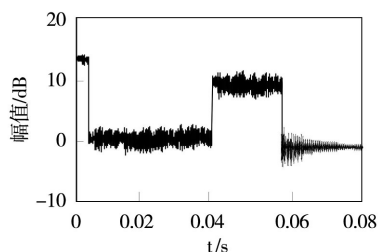


图2 原始系统数据分布情况示意图

以图2中的数据作为测试数据,对不同系统的筛选性能进行测试。在相同的测试环境下,对选择的三种系统分别进行测试,并将最终的结果进行对比与分析。

## 2.2 实验结果对比与分析

用图2中的数据作为输入数据,并利用本文所设计的数据筛选系统进行控制筛选,根据金融仓单中的可行数据语义的多样性和同义性进行知识数据的模糊处理,实

现风险数据的筛选,本文系统下得到的输出数据情况如图3所示。

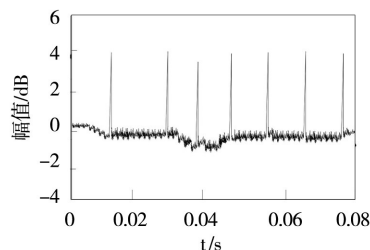


图3 系统输出数据

根据图3中的系统输出数据可以看出,本文设计的风控数据筛选系统输出的数据具有很强的特征聚类性。按照图中的输出结果对数据进行筛选精度分析,得到的三种系统精度分析结果如图4所示。

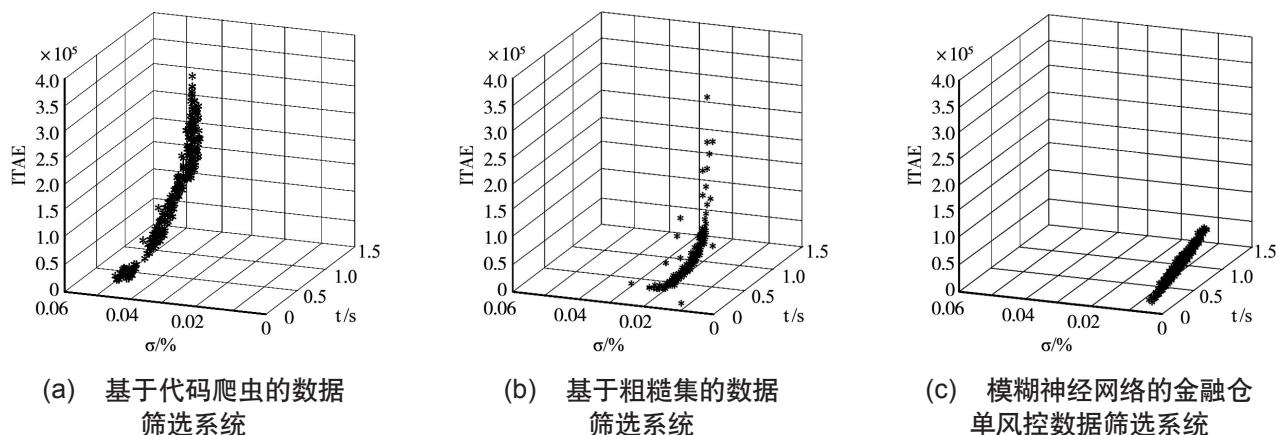


图4 不同筛选系统下的风险数据筛选准确性分析结果

将三个系统得到的结果利用第三方分析软件进行量化分析,能够求得各个系统筛选结果的置信度。经过多次测试,本文系统筛选出的风险数据置信度平均为0.948,基于代码爬虫的数据筛选系统的数据置信度平均为0.831,基于粗糙集的数据筛选系统的数据置信度平均为0.903。综上所述,本文系统筛选结果置信度最高,验证了设计系统的性能优越性。

## 3 结束语

金融风控的关键在于对数据的处理,针对金融风控的核心——仓单,需要有能够辅助决策的数据。目前国内银行和期货公司等金融机构所使用的数据,主要来源于仓库实际管理人员和工作人员。在实际应用中,对于不同仓单类型所需要的筛选标准存在一定差异,且不同仓库管理人员对数据处理的专业程度也存在差异。本文所设计的针对金融仓单风控数据筛选系统能够有效分析出各种类型的风险数据,在实时性方面具有较大的优

势。但是本文设计的系统还有一些待改进之处,在今后的设计研究中还要不断完善。

### 参考文献:

- [1] 雷达,常潇,刘子腾,等.基于DBSCAN聚类和数据筛选的系统谐波阻抗估算[J].电测与仪表,2022,59(1):93-98.
- [2] 邹文华,左帝,金开研.一种结合多重插补的ADS-B数据筛选算法[J].电讯技术,2021,61(4):461-467.
- [3] 王茂光,杨行.一种基于AP-Entropy选择集成的风控模型和算法[J].计算机科学,2021,48(S2):71-76,80.
- [4] 马彬,王双双,陈海波.基于区间二型模糊神经网络的垂直切换算法[J].电子学报,2021,49(5):928-935.
- [5] 刘多林,吕苗.基于网络爬虫结合关联大数据的用户信息提取[J].计算机仿真,2021,38(8):482-486.
- [6] 荆一楠,张寒冰,李智鑫,等.面向金融场景的下一代数据库测试基准研究[J].中国工程科学,2022,24(4):121-132.
- [7] 金磐石,李博涵,秦小麟,等.金融分布式数据库异步全局索引研究[J/OL].计算机科学与探索,2023,17(11):2784-2794.
- [8] 胡美慧,向志威.基于离散余弦变换的电力营销系统客户权

限自动识别方法[J]. 自动化技术与应用, 2022, 41(5): 125-129.

[9] 李正超, 梁志宏, 岳昆. 基于区块链的供应链金融数据存证模型[J]. 云南大学学报(自然科学版), 2022, 44(4): 681-687.

[10] 郭建平, 赵立龙. 基于前兆数据和经验模态分解的金融系统极值风险识别研究[J]. 统计与信息论坛, 2021, 36(3): 60-69.

[11] 甄志龙, 张居晓. 卡方统计中基于KL散度的高维文本数据特征筛选[J]. 统计与决策, 2022, 38(17): 43-46.

[12] 刘丽萍, 吕政. 基于高维高频金融数据的最小方差组合风险的估计及应用[J]. 系统科学与数学, 2021, 41(10): 2948-2964.

作者简介: 顾凌云(1976-), 男, 博士, 高级工程师, 研究方向: 人工智能, 深度学习, 金融科技, 自然语言处理, 语音识别, 信号处理.

(上接第73页)

信息评估结果为“中风险”, H9公司的财务风险较高, 达到“超高风险”等级; 2020年H5公司的财务风险加剧, 上升到“高风险”级别, 而H9公司的财务风险降低了一个级别; H1公司存在财务风险隐患, 其评分结果下降, 趋近中低风险临界值, 需加强风险防范意识。2021年, H1、H5公司的财务状况继续恶化, 分别升至“中风险”“超高风险”级别, 而H9公司的财务风险信息评估得分升高, 但仍未脱离“高风险”级别, 其余子公司的财务状况较好, 均处于低风险级别。实验结果表明, 本系统能够实现财务风险信息的自动识别。

## 2.5 财务风险信息识别系统的效率分析

响应效率与系统功耗水平是评价系统性能优劣的重要指标, 通过对不同任务处理量下响应时间、平均功耗分析, 验证本文系统的实际应用性能, 实验结果如图5所示。分析图5可知, 采用本文系统对不同数量的财务风险识别任务进行处理, 系统响应时间随着任务数的不断增加而呈增加趋势变化, 但增长幅度较为缓慢, 处理8 000个任务时, 响应时间仅为4 s左右。随着任务数的增长, 系统功耗基本维持在较低水平。实验结果表明, 本系统具有较高的响应效率, 且功耗较低, 系统性能突出。

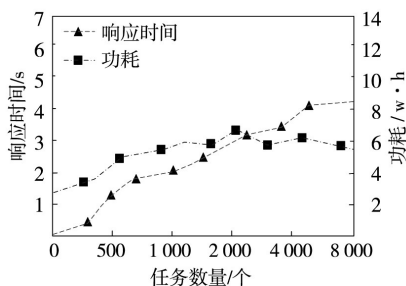


图5 本系统的运行性能分析

## 3 结束语

以某大型集团公司作为研究对象, 采用本文系统对该企业财务风险信息进行识别, 通过分析财务风险识别结果及系统的响应时间与功耗验证本系统的应用性能。实验结果表明本系统可确定财务风险信息指标, 识别财务风险信息等级, 本系统响应效率高、功耗小。

## 参考文献:

- [1] 严良, 李淑雯, 蒋梦婷, 等. 基于PCA的DE-SVM资源型企业财务风险识别模式研究[J]. 会计之友, 2019, 13(7): 58-65.
- [2] 张春梅, 赵明清, 官俊琪. 基于Lasso+SVM的制造业上市公司财务风险组合预警模型[J]. 数学的实践与认识, 2021, 51(5): 1-12.
- [3] 王小燕, 姚佳含. 基于聚类分析的惩罚约束财务风险预警模型[J]. 统计与决策, 2020, 36(2): 153-156.
- [4] 夏诗颖, 孔昭君, 丁宣宜, 等. 基于主成分分析和BP神经网络算法的股权众筹风险评价研究[J]. 技术经济, 2021, 40(11): 146-154.
- [5] 孙秀梅, 张敬云, 冯金红. 基于监测信息的医院财务风险监测预警方法[J]. 自动化技术与应用, 2022, 41(4): 132-135, 138.
- [6] 李勤敏, 郭进利. 基于主成分分析和神经网络对作者影响力的评估[J]. 情报学报, 2019, 38(7): 709-715.
- [7] 唐衍军, 黄益, 蒋翠珍. 区块链技术下的医院财务管理平台建设[J]. 卫生经济研究, 2020, 37(7): 62-65.
- [8] 林丹梅, 李珊珊, 肖世龙, 等. 基于长短期记忆神经网络的企业财务风险预警模型研究[J]. 南京理工大学学报, 2021, 45(3): 361-365, 374.
- [9] 金红美. ERP系统在企业财务管理中的应用[J]. 纳税, 2021, 15(30): 60-62.
- [10] 丁心怡, 吉久明. 基于系统动力学的企业财务风险情报分析模型[J]. 中国注册会计师, 2022, 11(7): 103-108.
- [11] 梁龙跃, 刘波. 基于文本挖掘的上市公司财务风险预警研究[J]. 计算机工程与应用, 2022, 58(4): 255-266.
- [12] 陈磊, 吴晓晖. 基于Hadoop的分布式集群大数据动态存储系统设计[J]. 中国电子科学研究院学报, 2019, 14(6): 593-598.
- [13] 郝玉锴, 吴姣, 李向东, 等. 一种机载高安全高可靠数据存储模块的设计[J]. 电光与控制, 2019, 26(1): 109-113.
- [14] 刘冰, 张睿全, 李亚群, 等. 基于UWB技术的无线透明传输模块组设计[J]. 中国电子科学研究院学报, 2019, 14(2): 168-176.

作者简介: 王澜斐(1992-), 女, 硕士, 助理会计师, 研究方向: 财务管理。

通信作者: 杨怡雯(1993-), 女, 本科, 助理会计师, 研究方向: 财务管理。