

数据分析与知识发现
Data Analysis and Knowledge Discovery
ISSN 2096-3467, CN 10-1478/G2

《数据分析与知识发现》网络首发论文

题目：融合知识图谱与人工免疫的企业风险识别模型研究
作者：李佳, 杨波, 杨美芳
网络首发日期：2024-01-19
引用格式：李佳, 杨波, 杨美芳. 融合知识图谱与人工免疫的企业风险识别模型研究 [J/OL]. 数据分析与知识发现.
<https://link.cnki.net/urlid/10.1478.G2.20240118.1802.004>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

融合知识图谱与人工免疫的企业风险识别模型研究

李 佳^{1,2}, 杨 波^{1,2}, 杨美芳^{1,2}

(1. 江西财经大学信息管理学院, 南昌 330032; 2. 江西财经大学信息资源管理研究所, 南昌 330032)

摘要: [目的] 提高企业风险识别的准确率, 降低潜在风险对企业造成的损失。 [方法] 提出一种融合知识图谱与人工免疫的风险识别模型。该模型利用知识抽取技术从文本信息中挖掘风险领域知识, 实现企业风险领域知识图谱的构建; 并通过企业风险事件描述文本信息与知识图谱中的风险实体进行实体链接, 获取更强的风险特征; 在此基础上, 运用人工免疫方法进行企业风险识别。 [结果] 该模型对企业风险识别率为 89%, 相对于基于神经网络的风险识别模型的识别率提升了 19%。 [局限] 本研究仅分析了企业年报中披露风险内容与企业新闻报道中的定性文本信息, 未来将尝试引入多维度与更完备的定量与定性信息来提高企业风险识别的准确率。 [结论] 本文提出的模型能够高效的关联企业内外部实时数据并开展企业风险分析, 为企业风险预控提供重要参考。

关键词: 风险识别; 知识图谱; 知识抽取; 人工免疫

分类号: TP391

Research on Risk Identification Model Combining Knowledge Graph and Artificial Immune

Li Jia^{1,2}, Yang Bo^{1,2}, Yang Meifang^{1,2}

(1. School of Information Management, Jiangxi University of Finance and Economics, Nanchang 330032; 2. Institute of Information Resource Management, Jiangxi University of Finance and Economics, Nanchang 330032)

Abstract: [Objective] To improve the accuracy of enterprise risk identification and reduce the loss caused by potential risks. [Methods] A risk recognition model combining knowledge graph and artificial immunity was proposed. The model uses knowledge extraction technology to mine risk domain knowledge from text information, and realizes the construction of enterprise risk domain knowledge graph. Through entity link between enterprise risk event description text information and risk entities in knowledge graph, stronger risk characteristics can be obtained. On this basis, the artificial immune method is used to identify enterprise risk. [Results] The risk recognition rate of the model was 89%, which was 19% higher than that of the neural network based risk recognition model. [Limitation] This study only analyzed the qualitative text information of risk disclosure in corporate annual reports and corporate news reports. In the future, we will try to introduce multi-dimensional and more complete quantitative and qualitative information to improve the accuracy of corporate risk identification. [Conclusion] The model proposed in this paper can efficiently correlate internal and external real-time data and carry out enterprise risk analysis, providing an important reference for enterprise risk pre-control.

Key words: risk identification; knowledge graph; knowledge extraction; artificial immunity

1 引言

近年来企业风险事件频发,涉事企业和风险损失不断攀升,从沸沸扬扬的共享单车相继倒闭事件,到康美药业财务造假事件,再到新冠疫情期旅游业、餐饮业与娱乐业等爆发的经营危机,各类危机事件或成为“新常态”,暴露了我国部分企业各类风险早已暗藏凶机^[1]。由于企业内外部环境复杂多变,各种风险因子高度集中,相关行业的部分企业受到市场成熟度、融资能力、产品推广能力、客户需求及互联网技术发展等因素的制约,面临巨大的复杂多样、突显性与动态性特征的风险^[2]。企业风险给投资者、供应商、客户和员工带来巨大的损失,逐渐成为影响其生存与发展的主要因素。因此,风险识别是企业顺利发展的关键环节,在企业风险管理与风险决策支持中扮演着重要的角色。

近几年,国内外学者对风险识别问题进行了相关研究,主要集中在风险识别指标选取与度量以及风险识别模型的研究。在风险识别指标选取与度量方面,典型的方法主要有文献归纳法^[3]、德尔菲法^[4]、层次分析法^[5]、风险清单法^[6]等。但这些方法多为简单的静态方法,难以衡量动态复杂的企业风险。在风险识别模型研究方面,大量学者采用支持向量机^[7]、神经网络^[8]和深度学习^[9]等方法进行风险识别。此类方法可减少模型对统计特征的依赖,同时能够挖掘领域深层次的风险特征,进一步提升模型的识别性能。然而这些方法一方面过于依赖于大规模的领域语料库,难以适应企业风险领域初期训练语料不足的情况,另一方面缺乏可解释性与透明性,难以在企业付诸实际应用。针对现有风险识别指标选取与度量研究的不足,有学者开始探讨了基于文本特征的财务风险预测问题^[10]。在现有文本特征提取的基础上,本文试图运用知识图谱中文本表示学习与知识抽取技术深入挖掘企业风险领域本体特征库,为风险识别模型提供底层完备的数据支撑。

随着多学科交叉融合理念的提出,现有的风险识别模型与方法也在不断地融合。一些学者从生命科学领域研究中的生物免疫系统得到启发,探究企业应对内外部环境变化时所形成的企业风险免疫问题^[11]。免疫系统经过生物进化被证明是有效的、适应性良好的动态防御系统^[12]。就免疫机理而言,生物免疫理论中外来入侵异己与风险管理理论中的企业内外部风险有着及其相似的特征,免疫系统防御异己的过程就是企业风险识别与处理的过程^[13]。借鉴生物免疫理论这一研究视角,本文拟将免疫系统作为参考体系,探索企业风险识别问题。针对风险识别模型研究的不足,我们将企业风险视为抗原,风险识别器视为抗体,运用已提取的抗原特征(风险领域本体特征库)以及少量的训练样本(风险领域语料库)生成初始抗体(初始风险识别器),并通过基因的克隆、交叉、变异与优选机制,不断淘汰与抗原“匹配度”低的抗体,补充新的“匹配度”更高的抗体,直至获得进化后最优的抗体。

鉴于此,本文运用知识图谱理论与方法提取企业风险领域本体特征库,并模拟生物免疫系统的运行机理,构建融合知识图谱与人工免疫的企业风险识别模型。该模型是不同于传统风险识别技术方法的有益尝试,为探究企业风险事件的免疫反应与风险识别提供了新思路。

2 企业风险识别相关研究

2.1 风险识别指标

企业风险爆发前,具有长期的潜伏性和隐蔽性,只有在量变积累到一定程度

才会带来质变，进而可能引发财务危机^[14]。因此，诸多经典的风控模型均将企业的财务指标作为衡量风险的重要特征。刘磊等人将资产负债率的波动作为反映金融风险变化的指标纳入风险预测模型中^[15]。MacKenzie 等人在前人研究的基础上，运用概率分布的数值衡量风险指标，将风险水平转换为具体的数字或分类指数^[16]。虽然目前企业风险识别指标体系正在不断完善，但是企业风险识别指标本身的度量优化却一度停止，仍主要使用财务相关的存量指标。这些存量指标存在固有的缺陷，大都基于企业过去表现的历史经验信息，忽略了企业风险识别相关的实时文本信息，难以描述企业其他方面的重要信息。因此，有学者开始尝试使用更多维度的数据指标进行风险识别，如企业审计数据、生产经营数据、证券交易数据等。

近年来，有研究表明定性领域文本能有效提升企业风险识别能力。如，企业年报、政策文件、重大事故描述、新闻事件等文本信息中蕴含着大量企业当前的经营状况以及未来可能面临的风险信息。李成刚等人通过提取企业年报中“管理层讨论与分析”的文本特征提升企业信用风险预警能力^[17]。李解等人借助 R 语言与文本挖掘方法提取地铁施工安全事故描述文本中的风险特征，为地铁施工安全风险识别提供重要参考^[18]。Liu 等人从互联网中提取有关我国企业风险相关的文本信息，构建用于度量我国企业风险水平的风险指数模型，能够准确识别近年来我国金融市场可能面临的风险^[19]。梁龙跃等人对上市公司年报中的描述性文本信息进行特征提取并应用于财务风险预警模型^[20]。

虽然上述研究者尝试将定性文本信息应用于风险识别与预警模型中，但风险领域的文本指标选取较少，仅根据少量的文本指标对企业风险进行识别，显然是较为片面的。知识图谱的出现为风险领域文本特征抽取提供了技术支撑。知识图谱的本质是揭示实体间关系的语义网络，旨在描述领域实体属性以及实体间关系^[23]。风险领域知识图谱能够围绕风险事件，从涉事主体、时间、地点、风险源、影响范围等角度组织风险知识，为动态知识表示、风险事件演化分析提供可行方法，也为企业风险特征提取提供了思路。因此，本文在前期的基础上，运用知识图谱技术与方法抽取风险实体特征作为企业风险识别指标，为风险识别模型提供更完备的数据支撑。

2.2 风险识别模型

风险识别模型的相关研究从基于统计方法的模型向基于神经网络的深度学习模型逐渐演进。早期的风险识别模型通常根据已有历史经验数据提取风险特征，并运用概率分析^[21]、因子分析^[22]与综合判别分析^[23]等方法进行风险识别。然而早期的风险识别模型需要较多的前提假设，且识别对象多为静态风险，无法充分考虑时间及不确定突发事件可能造成的风险演变或转移。

鉴于传统基于统计方法存在的缺陷，有学者采用基于神经网络的深度学习方法进行企业风险识别^[24]。Alawad 等人针对铁路客运系统风险的动态性和复杂性，提出基于卷积神经网络（CNN）的铁道风险事件识别模型，该模型有助于提高铁路系统风险预测的准确性，有效减少铁道风险事故的发生^[25]。Chen 等人运用我国石油市场数据训练长短期记忆网络（LSTM）模型，并基于 LSTM 模型对石油进口风险进行预测^[26]。与早期的方法相比，基于神经网络的深度学习方法无需对数据设定前提假设，并具有在大规模数据集上识别非线性关系的能力。然而，基于神经网络的深度学习模型过于依赖于大规模的领域语料库，难以适应企业风险领域初期训练语料不足的情况，而且当前提出的风险识别模型大多属于缺乏可解释性与透明性的黑箱模型，难以在企业风险管理中付诸实际应用。

随着多学科交叉融合理念的提出,多学科综合集成研究企业风险识别问题成为学科前沿。一些学者从生命科学领域研究中的生物免疫系统得到启发,探究社会领域的管理问题。国内学者王以华等人最早借鉴生物免疫系统理论研究企业组织的免疫力问题。他们认为,企业面临的内外部风险可类比为生物机体入侵的病原体,企业组织的免疫力能够识别内外部风险,并做出相应的免疫反应^[27]。吕萍等人在此基础上深入分析了组织免疫行为的动态机制,并探讨企业在复杂动态环境中如何提高自身组织免疫力以增强企业抗风险能力^[28]。国外学者 Haldane 等人在 Nature 上发文,运用免疫系统理论研究风险管理问题,提出金融风险管理的“免疫”思想^[29]。受生物免疫系统启发, Farmer 提出的人工免疫动态模型提供了强大的信息处理和问题求解范式,以简单的方法成功解决了大量的非线性动态问题。当前,人工免疫方法及其应用日益成熟,并应用于多个领域。杨青和刘星星等在总结借鉴免疫学、多 Agent 建模和仿真、应急管理理论和方法的基础上,建立了基于免疫学的应急预案多智能系统理论框架以及风险识别模型^[30]。杨超等人基于人工免疫理论与方法进行微博水军用户的识别,并使用新浪微博用户的真实数据验证了该方法的有效性^[31]。胡立伟等人构建了基于人工免疫机制的营运货车运行风险评价模型,为营运火车运行风险管理提供重要参考^[32]。

相对基于神经网络的深度学习方法,人工免疫方法初期无需大量的领域训练预料,还可以借助免疫识别、克隆选择、遗传进化等机理解决数据聚类与模式识别问题。因此,人工免疫方法较适合用于求解企业风险识别问题。在自然进化条件下,生物免疫系统是一个稳健可靠的防御系统。免疫系统抗击抗原和企业抵御风险的过程极为相似。抗原对免疫系统造成的危险,正如企业内外部风险给企业所带来的威胁^[36]。抗原代表企业风险,是风险识别的对象;抗体代表风险识别器。风险识别器根据企业已知的风险进行初始化,并通过克隆选择作用,不断淘汰“亲和度”低的,补充进新的“亲和度”更高的,直至获得进化的最优的风险识别器^[33]。人工免疫理论的发展和人工免疫应用领域的推广,为构建基于人工免疫的企业风险识别模型提供了可能。

3 融合知识图谱与人工免疫的风险识别方法

3.1 总体架构

为了更好的支持企业各类利益主体的风险识别,并辅助风险决策者制定合理的应对方案,本文从知识工程的视角出发,提出融合知识图谱与人工免疫的企业风险识别两阶段模型,如图 1 所示。

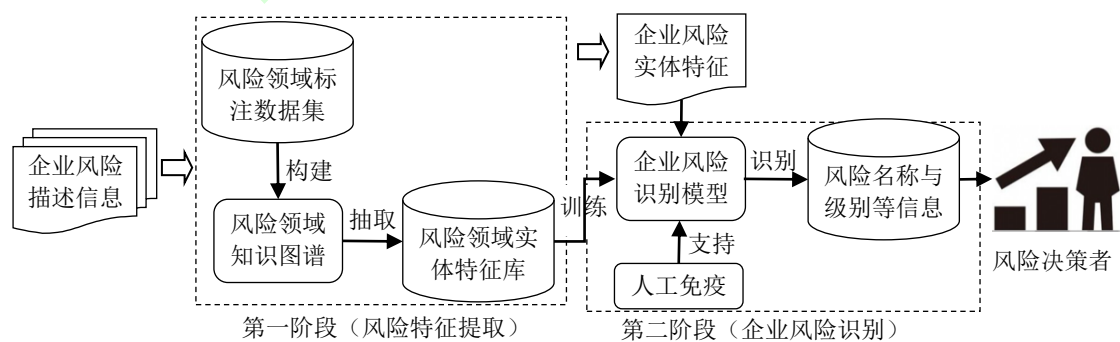


图 1 融合知识图谱与人工免疫的企业风险识别模型

Fig. 1 Enterprise risk identification model combining knowledge graph and artificial immunity

3.2 基于知识图谱的企业风险特征提取

(1) 企业风险领域知识图谱构建

知识图谱本质是一种语义网络，描述的是基于图的数据结构，由节点（实体）和边（实体间关系）组成^[33]。风险领域知识图谱以图结构反映风险领域的基本概念、实体及其关系，具备强大的语义处理与开放互联能力，能够更好的组织和管理海量的风险领域知识。本文采用半自动的方式构建面向企业风险领域实体抽取的知识图谱，以“企业-事件-场景”三大要素为核心实体（实体范畴如表 1 所示），利用企业风险领域数据的采集、处理与抽取实现对核心实体的多维扩展及实体关系的定义，提供企业风险领域实体的感知能力。

表 1 风险领域核心实体范畴描述

Tab. 1 Enterprise risk domain concept map

风险概念	风险实体	实体类型描述	示例
企业事件	风险管理组织	企业、政府、组织机构等	监管机构、政府、上市公司等
	具体的风险事件	国资委发布的企业风险事件	企业声誉风险、资金安全风险等
	发生时间	风险事件发生的时间	-
场景	发生地点	风险事件发生的地点	-
	风险源	引发风险的内在要素	企业负面新闻、新冠疫情等
	风险后果	某事件对企业影响的结果	客户满意度下降、GDP 下降等
	应对措施	解决风险问题采取的措施	加强员工培训、监控可疑贷款等

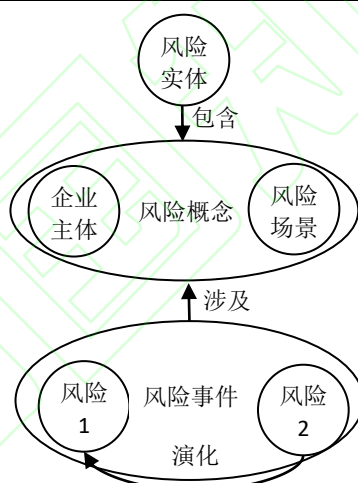


图 2 企业风险领域概念图谱

Fig. 2 Enterprise risk domain concept graph

企业风险领域知识图谱的构建采用自顶向下的模式，首先为知识图谱设计基本的概念模式，然后将从多源数据中抽取的实体加入知识库。企业风险领域概念图谱设计如图 2 所示。

(2) 企业风险领域实体识别

本文研究的企业风险领域实体识别主要包括风险场景实体识别与企业主体实体识别。风险场景实体识别主要从风险事件描述文本中识别出风险发生时间、发生地点、风险源、风险后果等场景实体，其目的是将口语化的风险事件描述与风险领域实体描述进行对应。企业主体实体识别主要是从风险事件描述文本中识别出涉事主体及其所属行业等实体，其目的是为后续解决风险场景相似而涉事主体不一致时的风险识别问题。

企业风险领域实体识别能够有效地将风险事件描述概念化，为风险特征体系构建提供统一的领域知识支持，从而使模型更便捷地进行风险识别。企业风险领域实体识别是指从风险领域文本中发现可能是风险场景实体或企业主体实体候选关键词。具体的，企业主体及其所属行业大都是名词短语，如“通信行业投资放缓”中的“通信行业”；企业风险后果大都是“名词+动词”形式的短语，如“公司营业收入下滑”中的“收入下滑”；企业风险措施大都是“动词+名词”形式的短语，如“关闭不安全生产线”等。因此，我们首先根据词性标注方法获取符合相应规则的候选关键词，然后在所有候选关键词中筛选风险场景实体和企业主体实体。

(3) 企业风险特征体系构建

基于已有领域知识图谱的研究以及风险领域核心实体的特点，本文从风险领域核心实体的文本向量表示和语言特征两个维度构建企业风险识别的特征体系，具体特征描述如表 2 所示。

表 2 企业风险识别模型的特征体系

Tab. 2 Characteristic system of enterprise risk identification model

类型	特征
文本向量表示	笔画 ELMo 嵌入预训练语言模型
语言特征	语法特性：代词、冠词、疑问词、数字等
	心理特性：认知词、情感词、感官词等
	其他：总结性词、标点符号、文本长度等

3.3 基于人工免疫的风险识别模型

本文构建了基于人工免疫的企业风险识别模型，如图 3 所示。该模型以 3.2 节所提取的风险领域实体特征库为基础，运用人工免疫方法构造风险识别器，并通过识别器对领域文本信息进行风险识别。该模型的关键技术主要包括风险识别器的生成与待测领域文本风险的识别。

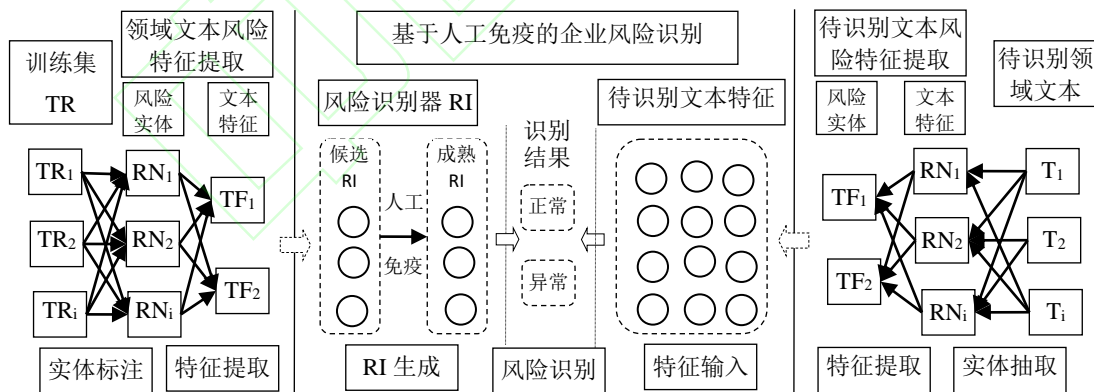


图 3 基于人工免疫的企业风险识别模型

Fig. 3 Enterprise risk identification model based on artificial immunity

(1) 风险识别器生成

免疫识别方法大都是通过模拟生物免疫系统中免疫细胞耐受与成熟过程来识别自体或异物，其核心基础算法是否定选择算法（Negative Selection Algorithm, NSA）^[34]。免疫识别方法的应用是基于实值形态空间展开，实值向量的匹配规则表示为数值间的相似度，通常使用欧氏距离、隶属函数、曼哈顿距离等进行度量。然而，企业风险识别特征体系是基于风险领域核心实体的文本向量

表示和语言特征等多个维度，传统的度量方法容易产生维度灾难问题，严重影响风险领域核心实体属性的选取以及模型的运行效率。另外，实值识别器的高重叠与“黑洞”问题一直是影响免疫识别方法应用效果的关键问题。基于此，本文借鉴晏义威等人提出的改进 V-detector 算法，应用风险领域实体特征库初始化候选风险识别器，并运用人工免疫方法对其进行耐受与择优，使其逐渐覆盖更大的非自体区域，有效缓解实值识别器的“维度灾难”以及高重叠与“黑洞”问题^[35]。风险识别器生成流程如图 4 所示。

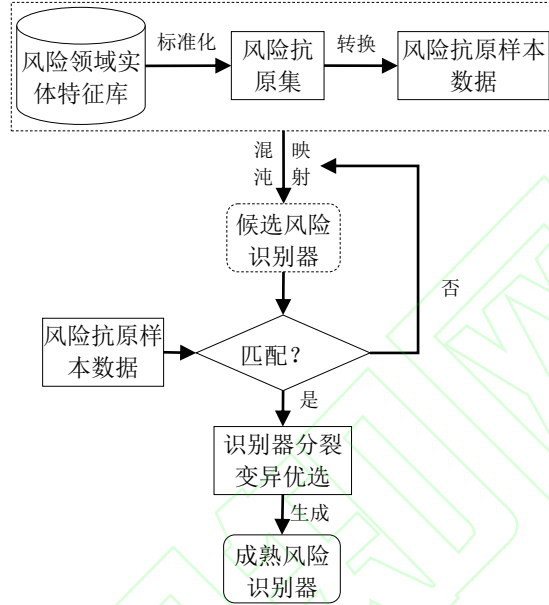


图 4 风险识别器生成流程

Fig. 4 Risk recognizer generation process

N 维空间下，自体（非风险抗原样本）和非自体（风险抗原样本）空间对应 R^n 的子集。抗原与风险识别器亲和力计算与匹配规则的设计如下。假设 AG_i 与 AB_j 分别为风险抗原集与风险识别器集。 $AG_i = [Ag_{i1}, Ag_{i2}, \dots, Ag_{in}]^T$ 为原始的已知风险抗原， Ag_{in} 描述的是第 i 个风险抗原在企业风险特征体系中的第 n 维特征向量。 $AB_j = [Ab_{j1}, Ab_{j2}, \dots, Ab_{jn}]^T$ 为初始化的候选风险识别器， Ab_{jn} 描述的是第 j 个风险识别器的第 n 维向量。 $T_{ij} = [t_i, t_1, t_2, \dots, t_n]^T$ 为第 i 个抗原与第 j 个风险识别器间的亲和度向量， t_n 为第 n 维风险特征的亲和度， t 为总亲和度， t_i 和 t 的计算分别为公式 (1) 和公式 (2)。

$$t_i = \sum_{i=1}^n \lambda_i \text{Aff}(Ag_m, Ab_m) \quad (1)$$

$$t = \sum_{i=1}^n t_i \quad (2)$$

其中， λ_i 为第 i 维风险特征所占的权重， $\text{Aff}(Ag_m, Ab_m)$ 为风险抗原与风险识别器（抗体）的第 n 维风险特征的亲和度，采用文献[36]提出的改进实值的曼哈顿距离进行计算， n 维向量 $v_g(g_{11}, g_{12}, \dots, g_{1n})$ 和 $v_b(b_{11}, b_{12}, \dots, b_{1n})$ 间的曼哈顿距离如公式 (3)。

$$d_{gb} = \sum_{i=1}^n |g_{1i} - b_{1i}| \quad (3)$$

N 维空间下亲和度匹配方法采用各维度风险特征进行独立匹配，削弱各维度风险特征间的耦合性，可有效解决传统度量方法的维度灾难问题，从根本上突破了维度属相的限制，无需进行降维即可高效运算。

根据改进的 V-detector 算法，本文将风险领域实体特征库转换为风险抗原样本数据，再应用风险抗原样本数据初始化候选风险识别器，初始的候选风险识别器的第 i 维属性从风险抗原样本数据的第 i 维特征中随机选取。风险抗原样本集与风险识别器集均采用“点+半径”的超球体模型表示。每个风险抗原样本设定固定的匹配半径 r_s ，反映了风险抗原覆盖区域。每个风险识别器定义为一个 n 维向量，并以二元组<识别中心点，识别半径>的形式进行表示。根据否定选择算法思想，若候选风险识别器未落在任何风险抗原样本半径 r_s 内（即候选风险识别器与风险抗原不匹配），则放弃该候选风险识别器；否则根据候选风险识别器的中心点 x 与风险抗原样本集的最短距离 d ，生成以 x 作为识别中心点，以 $d+r_s$ 为识别半径的候选风险识别器。

候选风险识别器通过与风险抗原样本的亲和度匹配后，将亲和度更高的候选风险识别器进行分裂变异优选，并将优选后的识别器加入到成熟风险识别器中。候选风险识别器分裂变异优选机制如下。

风险识别器分裂变异过程分为两步，若风险识别器的激活度达到阈值，则风险识别器分裂为两个子代抗体；然后对子代抗体以 P_m 概率进行变异操作。风险识别器的激活度由识别器浓度及其与抗原间亲和度决定，是评价识别器质量的指标。若风险识别器的浓度越低，且与抗原间亲和度越大，则风险识别器的激活度将越高，其对该类抗原的识别能力越强。本文借鉴文献[37]提出的免疫优化理论将风险识别器的激活度定义为公式（4）。

$$Atv(Ab_i) = Aff(Ab_i, Ag_i) \cdot e^{-\alpha \cdot Con(Ab_i)} \quad (4)$$

其中， $Atv(Ab_i)$ 为风险识别器的激活度， $Aff(Ab_i, Ag_i)$ 为风险识别器与抗原间的亲和度， α 为调节参数， $Con(Ab_i)$ 为风险识别器的浓度。风险识别器的浓度反映了识别器的多样性。若识别器浓度过高即识别器集合中存在大量相似的识别器，则识别范围将集中于局部区域的抗原集合，严重影响其识别效果。风险识别器的浓度定义为 $Con(Ab_i)$ ，计算公式如（5）。

$$Con(Ab_i) = \frac{\sum_{j=1}^{Num} sim(Ab_i, Ab_j)}{Num} \quad (5)$$

其中， Num 为风险识别器的总数， $sim(Ab_i, Ab_j)$ 为抗体 Ab_i 与抗体 Ab_j 间的相似度，可表示为公式（6）。

$$sim(Ab_i, Ab_j) = \begin{cases} 1, & Aff(Ab_i, Ab_j) < \delta \\ 0, & Aff(Ab_i, Ab_j) > \delta \end{cases} \quad (6)$$

其中， $Aff(Ab_i, Ab_j)$ 为抗体 i 与抗体 j 间的亲和度，反映了两个抗体间的相似程度，采用改进实值的曼哈顿距离进行计算； δ 为相似度阈值。

变异操作是使分裂后的子代识别器发生亲和度突变，产生更有潜力的新识别器，是免疫识别算法的关键，能进一步提升识别器的性能。基于实值编码的变异策略是在子代识别器中加入一个小扰动，使其稍微偏离原来的位置，落入原子代识别器的邻域空间。基于实值编码的变异策略可描述为公式（7）。

$$Mut(Ab_{i,j}) = \begin{cases} Ab_{i,j} + (Rd - 0.5) \cdot \xi, Rd < p_m \\ Ab_{i,j}, other \end{cases} \quad (7)$$

其中, $Ab_{i,j}$ 为抗体 Ab_i 的第 j 维, ξ 为自定义邻域范围, 可根据免疫优化过程自适应调整, Rd 为产生(0,1)间随机数的函数, p_m 为变异概率。变异机制在免疫识别模型中的作用在于保持抗体(识别器)的多样性, 低频度的变异可防止抗体(识别器)重要基因的丢失, 高频度的变异将使免疫识别模型接近随机检索。因此, 变异概率通常设置在[0.01-0.1]区间范围内^[38]。

风险识别器优选机制是对经过分裂变异后的识别器进行再选择, 剔除亲和度低的识别器, 保留亲和度更高的识别器进入成熟的风险识别器集。在优选过程中, 源风险识别器与分裂变异后的临时风险识别器共同组成新的集合, 免疫优选机制将保留此集合中亲和度最高的识别器, 放弃其他风险识别器。

(2) 待测领域文本风险识别

待测领域文本风险识别主要是利用上述生成的成熟风险识别器对待测领域文本进行风险识别, 其关键在于对待测领域文本的风险特征提取以及风险识别器与待测领域文本间的亲和力匹配。具体而言, 待测领域文本中包含了大量的风险领域实体特征, 首先通过 3.2 节提出的实体识别方法对其进行风险领域实体的识别, 再根据企业风险特征体系对待测领域文本进行风险领域实体特征提取, 获取企业风险领域实体特征集合 Ag_n , Ag_n 进入风险识别模型与成熟的风险识别器进行亲和力匹配。鉴于企业风险事件类别不同, 每一个特定类别的风险识别器对应着特定的亲和力匹配阈值 t , 最终亲和度高于匹配阈值 t 即为风险事件。识别出风险事件的识别器若达到其激活阈值, 将再次进行分裂变异优选, 筛选出更优秀的识别器更新到成熟风险识别器集。待测领域文本风险识别基本流程如图 5 所示。

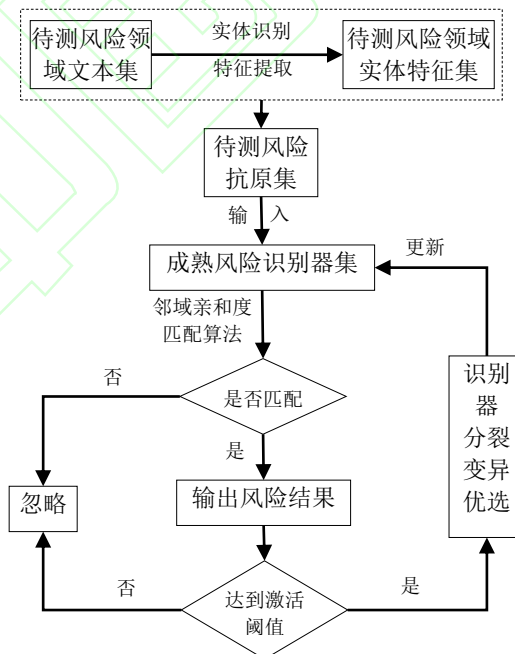


图 5 待测领域文本风险识别基本流程

Fig. 5 Basic process of text risk identification in the field to be tested

4 实验结果与分析

4.1 数据收集与预处理

本文研究数据来自三个部分，分别为学校购买的 DIB 风控系统中的企业风险案例数据 (sysData)、上市公司企业年报中披露的风险文本数据 (repData) 以及通过 python 爬虫爬取的互联网平台企业新闻数据 (recData)。其中，sysData 用以构建风险领域小规模知识图谱，辅助风险领域实体识别；repData 用以识别风险领域实体并构建风险领域实体特征库；recData 用以检验风险识别模型从未知待测文本中识别风险事件的能力。

根据研究数据功能的不同，本文运用 Mint 等人提出的远程监督算法对 sysData 数据进行风险领域实体、关系与风险事件提取；运用开源的文本标注工具 doccano 软件对 repData 数据与 recData 数据进行语义标注^[39]。repData 数据经过事件语义标注得到风险事件，再通过笔画 ELMo 预训练语言方法与 LIWC 工具分别对词向量与语言特征进行提取，构造有效的风险事件特征库，特征库中每一条记录包含着该风险事件在实体语义层面的特征。recData 数据经过事件语义标注后得到风险事件和非风险事件，并对这些事件进行相应的特征提取形成检验风险识别模型的数据集。

4.2 风险识别模型的实验分析

为验证本研究所提出的融合知识图谱与人工免疫的风险识别模型的有效性，本实验首先对风险领域实体识别性能进行测试，在此基础上训练基于人工免疫的风险识别模型，通过十折交叉验证模型的性能。

(1) 风险领域实体识别性能测试

本文以风险领域实体语义特征构建企业风险特征体系，并以此作为生成风险识别器的输入，直接影响风险识别模型的性能，而风险领域实体识别是实体语义特征提取的基础，对企业风险体系构建以及风险识别模型性能具有重要影响。因此，本实验首先测试风险领域实体识别模型的性能。

风险领域实体识别方法的主要创新点在于引入了风险领域小规模知识图谱特征，以此来解决风险领域特定实体识别所面临的领域知识要求较高，标注难度较大等问题。为验证该方法所做出的创新的有效性，本文将该方法与不引入知识图谱特征的识别方法进行比较。不引入知识图谱特征的识别方法通过文本分词后直接构造中文文本向量特征，不使用风险领域小规模知识图谱特征，其它步骤与本文提出的识别方法步骤一致。以上方法均使用 repData 数据集进行实验，并采用精度 (P)、召回率 (R) 与 F1 值作为测评指标。在使用相同的训练集和测试集的情况下，两种识别方法实验结果如表 3 所示。

表 3 不同实体识别方法的实验结果

Tab. 3 Experimental results of different entity recognition methods

实体识别方法	P	R	F1
融合知识图谱特征的实体识别方法	87.7%	91.5%	89.6%
不引入知识图谱特征的实体识别方法	78.8%	70.2%	74.3%

通过表 3 可以看出，本文提出的风险领域实体识别方法表现较优，F1 值能够达到 89.6%，证明模型引入知识图谱特征的合理性与有效性。引入知识图谱特征的实体识别效果得到显著提升的原因在于，原始风险领域描述文本语义较弱，部

分短语指代不清，而风险领域知识图谱中包含了领域特定实体的概念特征，这些概念特征能够使描述文本语义更加明确、部分短语指代更清晰，从而获取更好的实体识别效果。

(2) 基于人工免疫的风险识别性能测试

为衡量基于人工免疫的风险识别模型的性能，本文将从模型参数测定以及与其他模型性能对比两方面展开实验。

第一，模型参数测定实验

模型参数测定主要包括对匹配半径 r_s 、风险识别器激活度阈值 λ_{av} 以及变异概率 p_m 的取值测试。参数取值合适与否将直接影响识别器的生成效率、识别器的多样性以及模型最终的识别性能。本文采取风险识别研究中广泛使用的评价指标来衡量风险识别性能，这些指标主要包括识别率（IPR）与误判率（INR）^[40]。各项识别率的计算如公式（8）、公式（9）。

$$IPR = \frac{TP}{TP + FN} \quad (8)$$

$$INR = \frac{FP}{FP + TN} \quad (9)$$

其中，公式各变量使用混淆矩阵来定义，如表 4 所示。

表 4 风险识别混淆矩阵

Tab. 4 Risk identification confusion matrix

	实际情况（风险）	实际情况（非风险）
识别结果（风险）	TP	FP
识别结果（非风险）	FN	TN

首先，测试在不同匹配半径 r_s 情况下生成成熟风险识别器需要的时间。实验选取 repData 数据集中的风险数据作为风险抗原集训练成熟的风险识别器，将匹配半径 r_s 在区间[0, 0.1]上按照 0.01 步长进行依次取值，取 λ_{av} 为 0.006， p_m 为 0.02，统计生成 500 个成熟风险识别器所需的时间，实验结果如图 6 所示。从实验结果可得出，匹配半径 r_s 小于 0.1 时，成熟风险识别器生成时间较长；随着匹配半径 r_s 的增大，成熟风险识别器生成时间缩短，当匹配半径 r_s 在[0.01, 0.03]的范围内时，成熟风险识别器生成时间维持在较低的水平；当匹配半径 r_s 大于 0.26 时，成熟风险识别器生成速度开始下降。

其次，测试匹配半径 r_s 在[0.01, 0.03]的范围内成熟识别器的识别性能。实验随机抽取 recData 数据集中的样本组成测试数据集（其中包含 200 条风险事件和 100 条非风险事件），将匹配半径 r_s 在区间[0.01, 0.03]上按照 0.001 步长进行依次取值，取 λ_{av} 为 0.006， p_m 为 0.02，统计成熟风险识别器的性能，实验结果如图 7 所示。实验结果表明，当匹配半径 r_s 较小时，风险识别器的识别率与误判率均较高；随着匹配半径 r_s 的增大，风险识别器的识别率逐渐下降，误判率也维持在较低水平。当匹配半径 r_s 大于 0.02 时，风险识别器的识别率小于 0.6。综上实验分析，本文将选取匹配半径 r_s 为 0.016 进行后续实验。

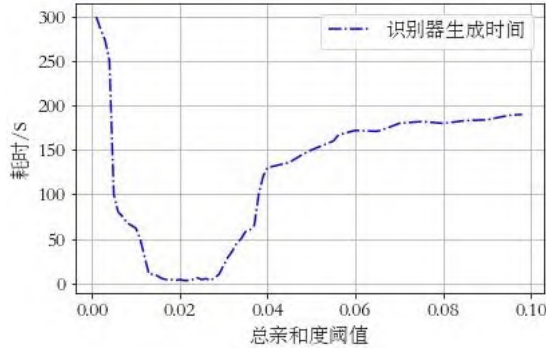


图 6 成熟风险识别器生成时间

Fig. 6 Maturity risk recognizer generation time

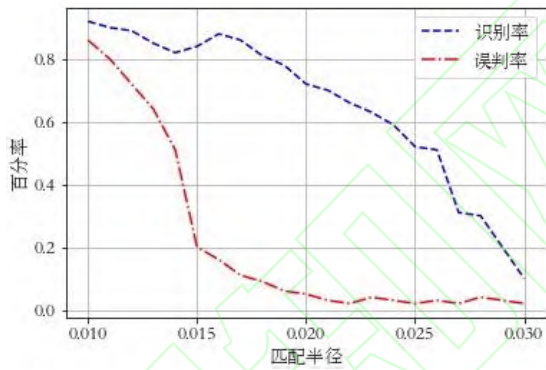


图 7 不同匹配半径成熟风险识别器识别性能

Fig. 7 Identification performance of mature risk recognizers with different matching radii

然后,测试风险识别器激活度阈值 λ_{av} 在 $[0, 0.1]$ 范围内成熟风险识别器的识别性能。实验随机抽取 recData 数据集中的样本组成测试数据集 (其中包含 200 条风险事件和 100 条非风险事件), 将 λ_{av} 在区间 $[0, 0.1]$ 上按照 0.001 步长进行依次取值, 取 r_s 为 0.016, p_m 为 0.02, 统计成熟风险识别器的性能, 实验结果如图 8 所示。实验结果表明, 随着激活度阈值 λ_{av} 的逐渐增大, 风险识别器的识别率呈上升趋势, 误判率也有所下降; 当激活度阈值大于 0.062 时, 风险识别器的识别率开始下降, 同时误判率也随之上升。因此, 本文将模型参数激活度阈值 λ_{av} 设置为 0.062。

最后, 测试变异概率 p_m 在 $[0.01, 0.1]$ 范围内成熟风险识别器的识别性能。实验随机抽取 recData 数据集中的样本组成测试数据集 (其中包含 200 条风险事件和 100 条非风险事件), 将 p_m 在区间 $[0.01, 0.1]$ 上按照 0.01 步长进行依次取值, 取 r_s 为 0.016, λ_{av} 为 0.062, 统计成熟风险识别器的性能, 实验结果如图 9 所示。实验结果表明, 随着变异概率的增加, 风险识别器识别率稍有上升, 误判率有所下降; 当变异概率增加到 0.05 时, 变异的风险识别器丧失原有优良的识别性能, 其识别率急剧下降, 误判率也加速上升。因此, 本文将模型参数变异概率 p_m 设置为 0.028。

综上实验结果表明, 当前实验数据与环境下, 模型参数匹配半径 r_s 、风险识别器激活度阈值 λ_{av} 以及变异概率 p_m 的取值分别为 0.016、0.062 与 0.028 时, 风险识别器的整体识别效果较好。

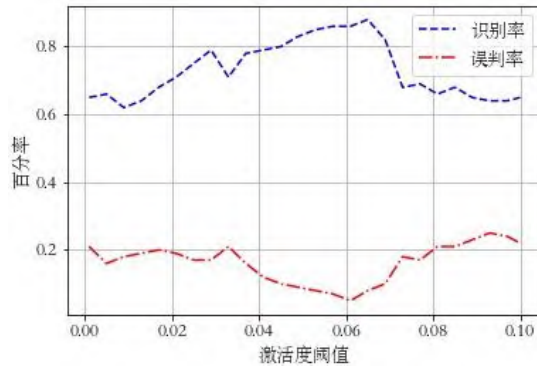


图 8 不同激活阈值成熟风险识别器识别性能

Fig. 8 Recognition performance of mature risk recognizers with different activation thresholds

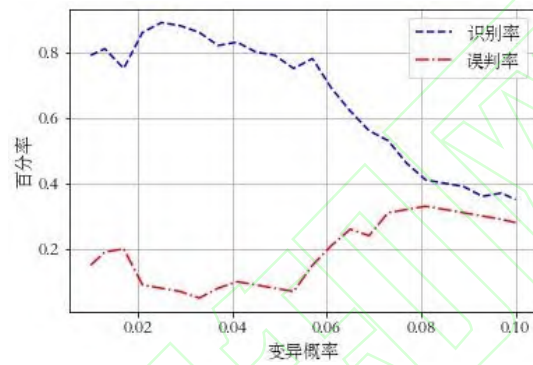


图 9 不同变异概率成熟风险识别器识别性能

Fig. 9 Identification performance of mature risk recognizer with different mutation probabilities

第二，待测领域文本风险识别实验

本文提出了基于知识图谱与人工免疫的风险识别模型（KGAIRI），其优势在于以风险领域实体特征代替文本特征，以此解决风险识别所面临的风险领域文本语料短缺、模型训练数据集不足的问题，并且在模型中应用风险抗原特征库生成候选风险识别器，进一步提升风险识别的准确率。为验证本文构建的风险识别模型的优越性，本文选用基于神经网络的风险识别模型（CNNRI）与基于传统人工免疫的风险识别模型（TAIRI）作为对照进行比较。基于神经网络的风险识别方法是直接使用风险领域文本进行风险分类。将现有风险领域语料分词，对分词结果计算词向量，并将其作为神经网络的输入进行风险识别。基于传统人工免疫的风险识别方法是采用随机方式初始化候选风险识别器，候选风险识别器经耐受择优后再进行风险识别。本实验随机抽取 recData 数据集中 5 组样本组成测试数据集（每组包含 200 条风险事件和 100 条非风险事件）。

在使用相同测试数据集的情况下，三种方法的识别率与误判率的对比实验结果如图 10 所示。由图可知，本文所提出的风险识别模型表现最优，识别率能够达到 0.89，误判率也维持在较低水平，表明 KGAIRI 模型使用风险领域实体特征表示风险特征的合理性以及应用风险特征库生成风险识别器的有效性；CNNRI 模型实验结果最差，主要原因在于相对于传统的文本训练数据集，本实验使用的风险领域数据集规模较小，一定程度上影响了 CNNRI 模型的训练效果。这也表明在风险领域数据集规模较小的情况下，KGAIRI 模型的优越性。另外，在同等实验环境下，KGAIRI 模型的识别效果优于 TAIRI 模型，原因在于 KGAIRI 模型应用风险特征库生成的风险识别器包含了更多的风险信息，能够覆盖更大的风险

抗原空间，从而有效提升了模型风险识别率。

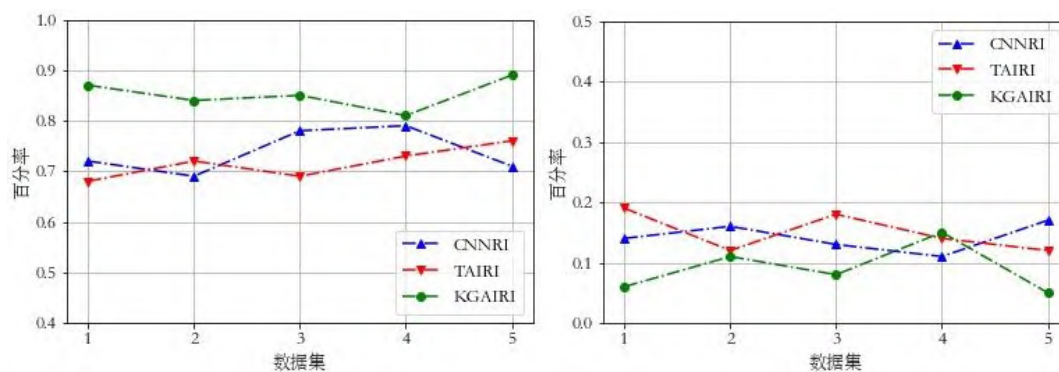


图 10 三种不同风险识别方法的性能比较

Fig. 10 Performance comparison of three different risk identification methods

5 结论

风险识别在企业风险管理与风险决策支持中扮演着重要的角色，成为目前应急管理领域核心研究课题之一。本研究面向高效获取企业风险指标体系的知识需求，针对风险领域初期语料库不足的问题，提出从知识抽取的角度获取风险领域实体特征的思想，融合知识图谱与人工免疫的方法构建企业风险识别模型，并以企业披露的风控数据与互联网爬取的企业新闻数据为例开展了实证分析，实现了知识工程视角下企业风险的高效识别。

本研究的实证结果在一定程度上验证了风险识别模型的有效性，能够较好地识别出信息服务企业当前面临的主要风险。此外，对于企业风险情报而言，本研究提出的模型能够高效的关联企业内外部实时数据并开展企业风险分析，是提升企业风险分析与风险预控时效性的有力手段，具有较高的实际应用价值。

然而，本研究仅分析了企业年报中披露风险内容与企业新闻报道中的定性文本信息，未来将尝试引入多维度与更完备的定量与定性信息来提高企业风险识别的准确率。另外，本研究提出的模型得到的实证分析结果缺乏鲁棒性，仅能作为风险管理层的辅助参考。如需准确深度的企业风险识别与分析，仍需资深风险管理专家的介入，利用领域专业知识确保风险识别的准确率，进而实现企业更高效的风险分析与防控。

参考文献

- [1] 李苍舒, 沈艳. 数字经济时代下新金融业态风险的识别、测度及防控[J]. 管理世界, 2019, 35(12): 53-69.(Li C S, Shen Y. Identification, measurement and prevention of risks of new financial forms in the era of digital economy[J]. Journal of Management World, 2019, 35(12): 53-69.)
- [2] Luo P., Hu Y. System risk evolution analysis and risk critical event identification based on event sequence diagram[J]. Reliability Engineering and System Safety, 2013, 114: 36-44.
- [3] Zhou Z P, Liu S, Qi H N. Mitigating subway construction collapse risk using Bayesian network modeling[J]. Automation in Construction, 2022, 143: 104-541.
- [4] 张鑫, 田雪灿, 刘鑫雅. 反复性视角下网络舆情风险评估指标体系研究[J]. 图书与情报, 2020(6): 123-135.(Zhang X, Tian X C, Liu X Y. Research on the Risk Assessment Index System of Internet Public Opinion from the Perspective of Recurrence[J]. Library & Information, 2020(6): 123-135.)
- [5] Kim S D. Characterization of unknown unknowns using separation principles in case study on Deepwater Horizon oil spill[J]. Journal of Risk Research, 2017, 20(1): 151-168.
- [6] Mojtahedi S M H, Mousavi S M, Makui A. Project risk identification and assessment simultaneously using multi-attribute group decision making technique[J]. Safety Science, 2010(48): 499-507.

- [7] Danenas P, Garsva G. Selection of support vector machines based classifiers for credit risk domain[J]. *Expert Systems with Applications*, 2015, 42(6): 3194-3204.
- [8] Wang F T, Ding L H, Yu H X, et al. Big data analytics on enterprise credit risk evaluation of e-Business platform[J]. *Information System and e-Business Management*, 2020, 18(3): 311-350.
- [9] Fu X L, Ouyang T X, Chen J P, et al. Listening to the investors: A novel framework for online lending default prediction using deep learning neural networks[J]. *Information Processing & Management*, 2020, 57(4): 1-13.
- [10] 陈艺云, 贺建风, 覃福东. 基于中文年报管理层讨论与分析文本特征的上市公司财务困境预测研究[J]. *预测*, 2018, 37(4): 53-59.(Chen Y Y, He J F, Tan F D. Financial Distress Prediction Based on the Textual Characteristics of Management Discussion & Analysis in the Chinese Annual Reports[J]. *Frontiers of Science and Technology of Engineering Management*, 2018, 37(4): 53-59.)
- [11] 吕萍, 王以华. 组织免疫行为和机制研究[J]. *管理学报*, 2009, 6(5): 607-614.(Lv P, Wang Y H. Study on the Behavior and Mechanism of Organizational Immunology[J]. *Chinese Journal of Management*, 2009, 6(5): 607-614.)
- [12] Matzinger P. The danger model: A renewed sense of self[J]. *Science*, 2002, 296(12): 301-305.
- [13] Yang Q, Ma H M. Multi-agent risk identifier model of emergency management system engineering based on immunology[J]. *Systems Engineering Procedia*, 2012, 4: 385-392.
- [14] Oreski S., Oreski G. Genetic algorithm-based heuristic for feature selection in credit risk assessment[J]. *Expert Systems with Applications*, 2014, 41(4): 2052-2064.
- [15] 刘磊, 刘健, 郭晓旭. 金融风险与风险传染——基于 CCA 方法的宏观金融网络分析[J]. *金融监管研究*, 2019, 9: 35-50.(Liu L, Liu J, Guo X X. Financial risk and Risk Contagion: Macro-financial network analysis based on CCA method[J]. *Financial Regulation Research*, 2019, 9: 35-50.)
- [16] MacKenzie, Cameron A. Summarizing Risk Using Risk Measures and Risk Indices[J]. *Risk Analysis*, 2014, 34(12): 2143-2162.
- [17] 李成刚, 贾鸿业, 赵光辉等. 基于信息披露文本的上市公司信用风险预警——来自中文年报管理层讨论与分析的经验证据[J]. *中国管理科学*, 2023, 31(2): 18-29.(Li C G, Jia H Y, Zhao G H, et al. Credit risk warning of listed companies based on information disclosure text: empirical evidence from management discussion and analysis of the Chinese annual report[J]. *Chinese Journal of Management Science*, 2023, 31(2): 18-29.)
- [18] 李解, 王建平, 许娜. 基于文本挖掘的地铁施工安全风险事故致险因素分析[J]. *隧道建设*, 2017, 37(2): 160-166.(Li J, Wang J P, Xu N. Analysis of safety risk factors for metro construction based on text mining method[J]. *Tunnel Construction*, 2017, 37(2): 160-166.)
- [19] Liu M Y, Li C Y, Wang S, et al. Digital transformation, risk-taking, and innovation: Evidence from data on listed enterprises in China[J]. *Journal of Innovation & Knowledge*, 2023, 8(1): 1-15.
- [20] 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.(Liang L Y, Liu B. Research on Financial Risk Early Warning of Listed Companies Based on Text Mining[J]. *Computer Engineering and Applications*, 2022, 58(4): 255-266.)
- [21] Bollacker K, Evans C, Paritosh P, et al. Freebase: A collaboratively created graph database for structuring human knowledge[C]// *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*. New York: ACM Press, 2008: 1247-1250
- [22] Li S C, Zhou Z Q, Li L P, et al. A new quantitative method for risk assessment of geological disasters in underground engineering: attribute interval evaluation theory (AIET)[J]. *Tunnelling and Underground Space Technology*, 2016, 53: 128-139.
- [23] Han Z, Huang S, Li H, et al. Risk assessment of digital library information security: a case study[J]. *The Electronic Library*, 2016, 34(3):471-487.
- [24] 陆静, 王捷. 基于贝叶斯网络的商业银行全面风险预警系统[J]. *系统工程理论与实践*, 2012, 32(2): 225-235.(Lu J, Wang J. An empirical study on risk factor identification and performance impact of enterprise IT

- projects[J]. *Statistics & Decision*, 2012, 32(2): 225-235.)
- [25] Alawad H, Kaewunruen S. An. A Deep Learning Approach Towards Railway Safety Risk Assessment[J]. *IEEE Access*, 2020, 8: 102811-102832.
- [26] Chen S., Song Y., Ding Y., et al. Using long short-term memory model to study risk assessment and prediction of China's oil import from the perspective of resilience theory[J]. *Energy*, 2021, 215.
- [27] 吕萍, 王以华. 基于组织免疫视角的企业适应性研究[J]. *科研管理*, 2008, 29(1): 164-171.(Lü P, Wang Y H. The study on organizational immunity based enterprise adaptation[J]. *Science Research Management*, 2008, 29(1): 164-171.)
- [28] 王以华, 吕萍, 徐波等. 组织免疫研究初探[J]. *科学性与科学技术管理*, 2006, 6: 133-139.(Wang Y H, Lü P, Xu B, et al. The study on organization immunity[J]. *Science of Science and Management of S.& T.*, 2006, 6: 133-139.)
- [29] Haldane A G, May R M. Systemic risk in banking ecosystems[J]. *Nature*, 2011, 469(7330): 351-355.
- [30] 杨青, 刘星星等. 基于免疫系统的非常规突发事件风险识别模型[J]. *管理科学学报*, 2015, 18(4): 50-59.(Yang Q, Liu X X, et al. Risk identification model of unconventional crisis based on immune system[J]. *Journal of Management Sciences in China*, 2015, 18(4): 50-59.)
- [31] 杨超, 秦廷栋, 范波等. 基于人工免疫危险理论的微博水军用户检测研究[J]. *计算机科学*, 2018, 45(11): 138-159.(Yang C, Qin T D, Fan B, et al. Study on detection of Weibo spammers based on danger theory in artificial immunity system[J]. *Computer Science*, 2018, 45(11): 138-159.)
- [32] 胡立伟, 何越人, 李耀平等. 基于人工免疫机制的营运火车运行风险评价研究[J]. *交通运输系统工程与信息*, 2021, 21(1): 149-155.(Hu L W, He Y R, Li Y P. Commercial truck driving risk evaluation based on artificial immune mechanism[J]. *Journal of Transportation Systems Engineering and Information Technology*, 2021, 21(1): 149-155.)
- [33] Wang Q, Ji Y D, Hao Y S, et al. GRL: Knowledge graph completion with GAN-based reinforcement learning[J]. *Knowledge-Based Systems*, 2020, 209: 1-8.
- [34] 朱方东, 李涛, 杨进. 基于免疫MHC的否定选择算法孔洞修复方法[J]. *中国科学: 信息科学*, 2020, 50(10): 1529-1543.(Zhu D F, Li T, Yang J. A negative selection algorithm hole improvement method based on MHC[J]. *Scientia Sinica(Informationis)*, 2020, 50(10): 1529-1543.)
- [35] 晏义威, 张凤斌, 杨晓君. 基于改进的可变半径检测器实值否定选择算法[J]. *哈尔滨理工大学学报*, 2009, 14(2): 13-16,20.(Yan Y W, Zhang F B, Yang X J. Based on extendent realvalued negative selection algorithm with variable sized detectors[J]. *Journal of Harbin University of Science and Technology*, 2009, 14(2): 13-16,20.)
- [36] 窦家维, 葛雪, 王颖因. 保护隐私的曼哈顿距离计算及其推广应用[J]. *计算机学报*, 2020, 43(2): 352-365.(Dou J W, Ge X, Wang Y N. Secure manhattan distance computation and its application[J]. *Chinese Journal of Computers*, 2020, 43(2): 352-365.)
- [37] 陶媛. 基于生物免疫的动态风险识别模型研究与应用[J]. 上海大学, 2011.(Tao Y. The research and application on the biological immunity based dynamic risk identification model[D]. Shanghai University, 2013.)
- [38] 包子阳, 余继周. 智能优化算法及其 MATLAB 实例[M]. 电子工业出版社, 2016.(Bao Z Y, Yu J Z. Intelligent optimization algorithm and MATLAB example[M]. Electronic Industry Press, 2016.)
- [39] Mints M, Bills S, Snow R, et al. Distant supervision for relation extraction without labeled data[C]//Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2. Singapore, 2009: 1003-1011.
- [40] 杨波. DRI 生物免疫视角扩展: 机理、方法与仿真研究[J]. *计算机工程与科学*, 2019, 41(3): 567-574.(Yang

B. DRI biological immune perspective extension: mechanism, method and simulation[J]. Computer Engineering & Science, 2019, 41(3): 567-574.)

(通讯作者: 杨美芳, ORCID: 0000-0002-4360-0183, E-mail: yangmeifang@jxufe.edu.cn.)

基金项目: 本文系国家自然科学基金项目“基于免疫方法的新创企业成长风险管理知识服务模式研究”(项目编号: 72064015)、江西省教育厅科技项目“江西省五大科技创新城市创新发展模式与路径研究”(项目编号: GJJ200536)和江西省社会科学“十三五”规划项目“面向新创企业成长风险管理的知识服务机制研究”(项目编号: 19TQ01)的研究成果之一。

This paper is supported by National Natural Science Foundation project (Grant No. 72064015) and the Science and Technology project of Jiangxi Provincial Department of Education (Grant No: GJJ200536) and the "13th Five-Year Plan" project of Jiangxi Province Social Sciences (Grant No. 19TQ01).

作者贡献说明:

李佳: 论文起草, 采集、清晰和分析数据。

杨波: 提出研究思路, 论文修改。

杨美芳: 设计研究方案, 论文修改。

利益冲突声明:

所有作者声明不存在利益冲突关系。