

大宗商品供应链风险识别

赵昕,戚丹阳,刘华琼*

山东交通学院交通与物流工程学院,山东 济南 250357

摘要:针对传统大宗商品供应链风险识别方法的识别角度不全面,识别结果准确度较低等问题,采用文本挖掘方法,建立包括数据采集及语料库构建、数据预处理、风险识别的大宗商品供应链风险识别模型框架,从中国知网、万方数据知识服务平台中获取大宗商品供应链管理相关研究文献,构建包含不同文献数的3个语料库,对语料库数据进行词频分析、N-gram分析、相关性分析、累计词频-信息熵(term frequency-information entropy, TF-H)降维及潜在狄利克雷分配(latent Dirichlet allocation, LDA)主题建模,并将风险识别结果与传统供应链风险识别方法的识别结果进行对比,验证方法的有效性。结果表明:通过LDA主题模型生成20个大宗商品供应链风险主题,每个主题从不同角度展示当前大宗商品供应链面临的风险,将识别出的大宗供应链风险分为市场风险、物流风险、金融风险、环境风险、管理风险、合作风险6种类型;文本挖掘方法与其他传统风险识别方法的风险识别结果具有较强的耦合度,且识别维度更全面、识别结果更准确。文本挖掘技术可全面、准确地识别供应链风险因素,可为大宗商品供应链风险识别提供理论支撑。

关键词:大宗商品;供应链管理;风险识别;文本挖掘;LDA

中图分类号:[U-9];TP391.1

文献标志码:A

文章编号:1672-0032(2025)01-0024-11

引用格式:赵昕,戚丹阳,刘华琼.大宗商品供应链风险识别[J].山东交通学院学报,2025,33(1):24-34.

ZHAO Xin, QI Danyang, LIU Huaqiong. Risk identification of bulk commodity supply chains[J]. Journal of Shandong Jiaotong University, 2025, 33(1): 24-34.

0 引言

大宗商品供应链具有交易规模大、涉及环节多、金融属性强、分布范围广、国际化程度高等特点,在经济全球化背景下,大宗商品供应链网络面临较大风险^[1]。党的二十大报告提出:“加快建设现代化经济体系,着力提高全要素生产率,着力提升产业链供应链韧性和安全水平”。《“十四五”现代物流发展规划》提出:“加快构建全球供应链物流服务网络,保持产业链供应链稳定”。研究大宗商品供应链风险管理,引导制造企业与物流企业建立互利共赢的长期战略合作关系,加快物流枢纽资源整合,构建国际国内物流大通道。

在研究大宗商品供应链特性方面,尹志^[2]采用向量自回归模型分析国际大宗商品价格波动对我国经济的影响,并提出应对策略;方先明等^[3]采用时变参数的向量自回归模型,检验地缘冲突风险对我国农产品期货价格冲击的时变特征;宋华等^[4]探索金融科技平台如何帮助金融机构提高对中小企业信用风险的评估能力,推动实现供应链金融创新模式。大宗商品供应链特性决定供应链风险类型复杂多样,易受内、外部环境影响,可根据大宗商品供应链特性识别风险因素。

作为供应链风险管理的起点,供应链风险识别是供应链风险评估和控制的前提。风险识别准确性和

收稿日期:2024-04-24

基金项目:山东省金融应用重点研究项目(2022JRZZ-24)

第一作者简介:赵昕(1998—),男,山东寿光人,硕士研究生,主要研究方向为交通运输,E-mail:571470023@qq.com。

*通信作者简介:刘华琼(1979—),男,安徽巢湖人,教授,硕士研究生导师,工学博士,主要研究方向为物流规划、物流优化、电商物流协同,E-mail:lhq5983442@163.com。

完整性直接影响风险评估和控制的效果和效率。传统的供应链风险识别方法包括文献分析法、故障树分析法、德尔菲法、供应链运作参考(supply-chain operations reference, SCOR)模型等。张敏等^[5]等分层归类农产品直播电商供应链风险因素,采用决策试验与评价实验法分析因素间的相互作用,通过ABC分类法识别关键风险因素;李国亮^[6]采用SCOR模型从计划、采购、生产、配送、退货等环节识别企业供应链风险;Cao等^[7]通过整合环境、经济和社会因素,提出可持续供应链运营综合决策模型,为平衡供应链运作中可持续发展目标与企业盈利目标提供决策工具和策略建议;谢乐^[8]以“农超对接”为研究对象,基于贝叶斯网络构建“农超对接”供应链风险预警模型,结合实证分析识别“农超对接”供应链的主要风险因素;Aqlan等^[9]结合专家知识、供应链结构、调查访问数据等识别供应链潜在风险。传统的供应链风险识别方法和数据来源存在一定局限性和主观性,由于数据多为人工采集和整理,限制了样本数和风险识别的角度。大宗商品贸易交易环节越复杂,交易工具种类越多,贸易业务风险表现形式越隐蔽,传统的风险识别方法易出现识别角度不全面、识别结果不准确等问题。

随着人工智能和大数据技术的发展,在风险识别中可采用多种方法挖掘分析海量数据,精准识别风险。Hassan^[10]通过机器学习方法识别汽车供应链风险,优化供应链管理;牛莉霞等^[11]通过大数据技术构建数据感知-数据分析-数据服务一体化的煤矿安全风险治理模式,提高煤矿风险识别效率和治理能力;Zhao等^[12]等通过网络爬虫技术爬取石油市场新闻数据,采用潜在狄利克雷分配(latent Dirichlet allocation, LDA)识别石油供应链的28个风险主题;刘心男等^[13]采用文本挖掘法分析建设工程质量检测机构检查报告,识别建设工程检测质量风险,检测工作合规风险。人工智能和大数据技术为供应链风险识别提供更科学的视角和工具,使风险管理从传统的经验判断和定性分析转向数据驱动决策。人工智能和大数据技术在风险识别领域已取得一定成果,但从大宗商品出发识别供应链风险的研究较少。

本文基于文本挖掘方法,采用词频分析、N-gram分析、相关性分析识别大宗商品供应链风险,将信息熵引入特征权重评价,克服词频-逆文档频率法(term frequency-inverse document frequency, TF-IDF)在提取特征时未考虑关键词出现频率和分布的问题,通过累计词频-信息熵(term frequency-information entropy, TF-H)方法对语料库降维,采用LDA实现主题建模,对大宗商品供应链风险进行识别与分类,以期为大宗商品供应链风险识别提供更科学的工具和方法。

1 大宗商品供应链风险识别模型框架

大宗商品供应链风险识别模型框架包括数据采集及语料库构建、数据预处理、风险识别三部分,具体框架如图1所示。

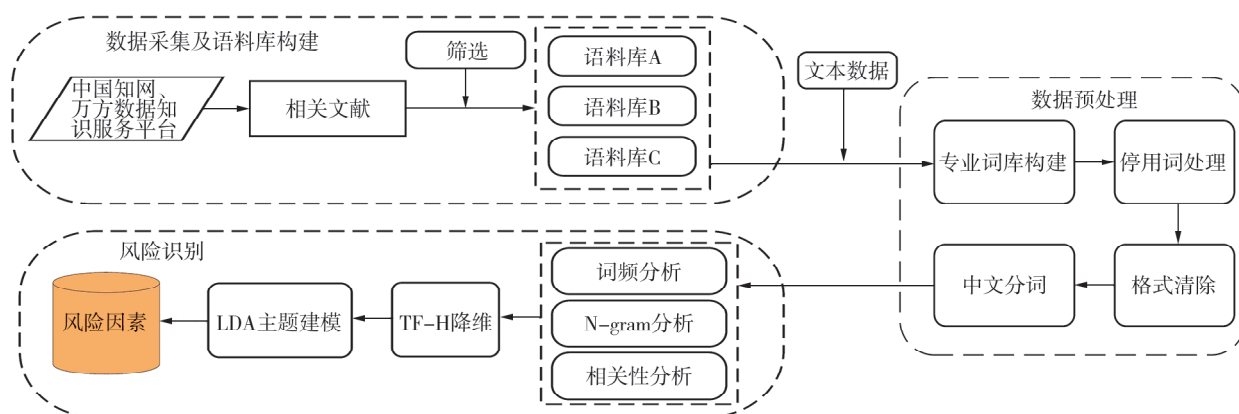


图1 风险识别模型框架

从中国知网、万方数据知识服务平台中检索与大宗商品供应链管理相关的文献构建语料库数据。通过2个数据库的高级检索功能,将检索主题限定为“大宗商品”和“供应链”,时间设定为2000—2023年,

人工筛除相关性较小的文献后,共得到 653 篇文献存入语料库 A。在数据挖掘领域,数据越多,结果越客观,但文本挖掘中的关联规则计算机制与传统的关联分析不同,相关性表示 2 个术语在同一文献中共同出现的频次,随文献数增大,共现频率可能会被稀释。将检索主题进一步限定为“大宗商品”“供应链”“风险”,时间不变,将语料库 A 进一步减少为包含 106 篇文献的语料库 B^[14]。

语料库 A、B 文献的基本属性如表 1~5 所示,仅统计语料库样本数据排名前 5 的属性类别。

表 1 语料库 A、B 文献来源类别				
语料库	文献篇数			
	北大核心	CSSCI	CSCD	AMI
A	75	53	6	45
B	59	22	1	21

表 2 语料库 A、B 文献主要发表年份							
语料库	文献篇数						
	2014 年	2017 年	2019 年	2020 年	2021 年	2022 年	2023 年
A	45			75	78	151	61
B		7	7	10	16	50	

表 3 语料库 A、B 文献主题						
语料库	文献篇数					
	产业链	供应链	供应链金融	大宗商品	大宗商品价格	大宗商品贸易
A		78	49	172	60	35
B	8	27	21	43	21	

表 4 语料库 A、B 文献主要学科来源							
语料库	文献篇数						
	企业经济	宏观经济管理与可持续发展	工业经济	经济体制改革	贸易经济	金融	
A	102	94	79		174	119	
B	15	4		10	16	33	

表 5 语料库 A、B 文献的主要来源期刊							
语料库	文献篇数						
	中国储运	中国外汇	中国物流与采购	全球流通经济	浙江经济	财经界	铁路采购与物流
A	46		47		7	7	8
B	8	4	3	5		2	

为获得更显著的相关性结果,通过充分阅读,人工筛选与大宗供应链风险主题高度相关的 12 篇文献存入语料库 C,其文献主题为大宗商品贸易、供应链风险,学科分布前 3 名分别为企业经济、工业经济、贸易经济。3 个语料库除了相关性有差异,分析结果均稳定。采用 Python 编程语言实现文本数据处理、结果分析和可视化。

2 大宗商品供应链风险识别方法

2.1 数据预处理

在风险识别前需进行数据预处理,对语料库调优,保证模型识别结果的准确性和全面性。数据预处理包括专业词库构建、格式清除、停用词处理和中文分词。

中文分词采用结巴分词库中的精准模式,但分词库缺乏对大宗商品供应链风险管理领域的知识,分词结果易出现将专业名词分割的情况,如,将“区块链”分为“区块”和“链”。构建专业词库实现精准分词,通过搜集整理搜狗细胞词库中的《物流词汇大全》《交通运输》《国际贸易名词》《外贸词汇大全》等多个相关领域词库,去除相同词作为专业词库^[15];格式清除可提高分词准确性和分词效率,清除文本中存在的英文、数字、标点符号、空格等;停用词处理是通过停用词库过滤文本中无意义的词^[16],停用词库包含《哈工大停用词表》及“摘要”“参考文献”等文献常见但无实际意义的词;最后通过结巴分词完成数据预处理。

2.2 词频及术语相关性分析

文献-术语矩阵是计算术语频率和术语间相关性的方法^[17]。语料库 C 中 12 篇文献按术语频率排序的文献-术语矩阵热力图如图 2 所示,图中矩阵行向量为语料库 C 中的文献,列向量为语料库 C 中出现频率排名前 10 位的术语。通过文献-术语矩阵分析可获得某术语在语料库和单篇文献中出现的频数,分析大宗供应链风险相关文献,推导重要的风险因素及相关术语。

术语间的相关性也是识别风险因素的重要方法之一,将某术语与术语“风险”共现的文献数除以包含术语“风险”的总文献数得到共现相关性得分,该得分反映某术语与术语“风险”间的关联强度,分数越高,与“风险”的关系越紧密。

N-gram 分析是自然语言处理中的常用方法^[18],通过分析和计量文本中连续的 N 个项(如单词或字符)序列的出现频率,揭示术语间的潜在关系;其中 Bi-gram 是将文本中相邻的 2 个词作为 1 个单元进行分析。本文采用 Bi-gram 分析进一步挖掘术语“风险”与其他术语间的联系。

2.3 特征提取及降维

在文本挖掘领域,特征是数据集中独立变量,TF-IDF 是使用最普遍的特征参数项统计方法^[19],若某术语在某文献中出现频率较大,在整个语料库出现频率较小时,认为该术语的重要性较高,但该方法在提取特征时未考虑术语分布问题,TF-H 算法将信息熵 H 引入特征权重评价中,可有效解决此问题^[20]。

假设术语 S_i 分布在 m 篇文献中, S_i 在语料库中的概率分布

$$p_i = T_j^i / \sum_{j=1}^m T_j^i,$$

式中 T_j^i 为术语 S_i 在第 j 篇文献中出现的频数。

术语 S_i 在文献中分布程度的信息熵 $H(S_i) = -\sum p_i \log p_i$ 。

术语 S_i 的 $H(S_i)$ 越大,随机性越大,风险因素发生的不确定性越大。结合词频和信息熵,术语 S_i 的词频-信息熵 $\omega(S_i) = T_j^i H(S_i) = -T_j^i \sum p_i \log p_i$ 。

文献较多时,特征参数的维度会过高,干扰后续分析,需对特征项降维处理。相较于传统的累计熵权词频和高频词界定公式等,累计 TF-H 法是将帕累托分析法的概念引入累计词频中,能更准确地剔除重要程度低的特征值。在该方法中,将所有特征项按词频-信息熵降序排列,计算每个特征项的词频-信息熵的累计百分比,保留累计百分比位 0~90% 的特征项,剔除其余特征项,减少维度,保留关键特征。特征项词频-信息熵分布如图 3 所示。

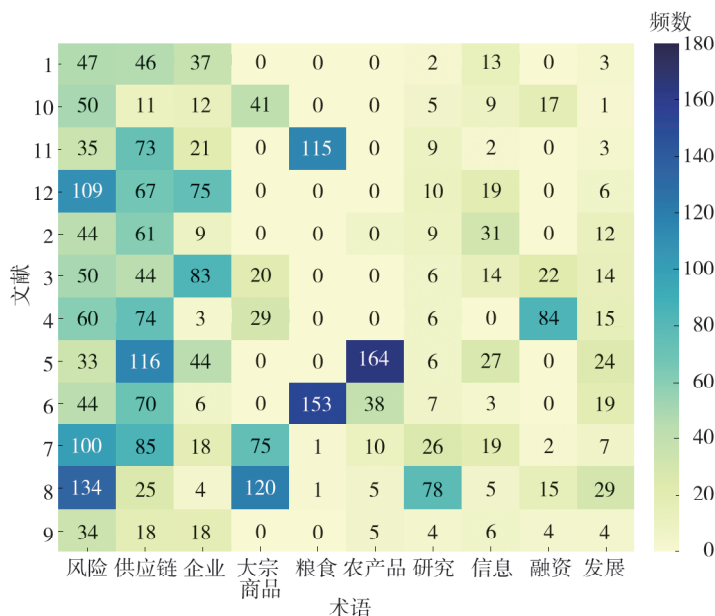


图2 按术语频率排序的文献-术语矩阵热力图

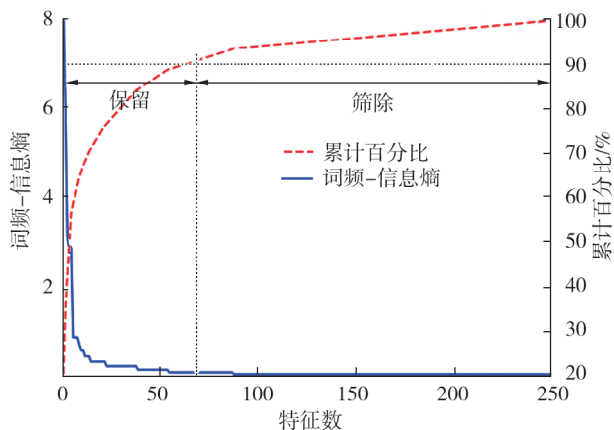


图3 特征项词频-信息熵分布

2.4 LDA 主题模型

LDA 是从文本数据中提取主题的概率生成模型^[21]。假设每个文献由多个主题组成,每个主题由词组的概率分布组成,通过推断文献中的主题分布和主题中的单词分布实现主题聚类,LDA 主题模型运行原理如图 4 所示。

对模型中的每篇文献,基于 α 选择 θ ,对文献中的每个词项从 θ 中选择 Z ,从与该主题相关联的 Φ 中选择 W ,重复以上过程,直到文献的所有词项分配主题。

LDA 主题模型需预设主题数,通过试验确定合适的主题数,常见方法是训练多个 LDA 主题模型,每个模型采用不同的主题数,计算每个 LDA 主题模型的主题困惑度^[22]。主题困惑度是衡量模型性能的重要指标之一,主题困惑度越小,模型的预测能力越好。比较不同模型的主题困惑度,确定最优主题数。主题困惑度

$$P(D_{\text{test}}) = \exp\left(-\sum_{d=1}^{m_{\text{test}}} \log P(W_d) / \sum_{d=1}^{m_{\text{test}}} N_d\right),$$

式中: D_{test} 为测试集文献集合; m_{test} 为测试文献数; W_d 为文献 d 中的词项集合; N_d 为文献 d 的词项总和; $P(W_d)$ 为文献 d 的词项集合在模型中的联合概率,给定主题分布后假设文献中的词项独立, $P(W_d)$ 可分解为文献中各词项概率之积, $P(W_d) = \prod_{n=1}^{N_d} P(W_{d,n})$,其中 $P(W_{d,n})$ 是文献 d 中第 n 个词项 $W_{d,n}$ 的概率。

通过混合文献的主题分布和主题的词分布计算 $P(W_{d,n})$,公式为:

$$P(W_{d,n}) = \sum_{i=1}^K P(Z_i | d) P(W_{d,n} | Z_i),$$

式中: K 为主题数, $P(Z_i | d)$ 为给定文献 d 时主题 Z_i 的条件概率, $P(W_{d,n} | Z_i)$ 是给定 Z_i 时词项 $W_{d,n}$ 的条件概率。

3 结果分析

3.1 大宗商品供应链风险识别

对 3 个语料库进行词频分析、相关性分析和 LDA 主题建模,基于文献关联规则机制,语料库术语越多,相关性越低,需缩减语料库规模。3 个语料库的概要如表 6 所示,经预处理后,语料库中的字符数和特征数均显著减少。

表 6 3 个语料库的概要

语料库	文献数	字符数	预处理后字符数	分词数	特征数	经 TF-H 处理后的特征数
A	653	3 992 740	2 532 445	759 033	19 699	7 599
B	106	931 267	626 881	219 411	20 411	5 194
C	12	86 593	60 273	17 684	2 825	343

3 个语料库的维恩图如图 5 所示。图 5 展示 3 个语料库词频最高的前 50 个词间的逻辑关系,语料库 A、B、C 各有 6、6、26 个独立元素;语料库 A 和 B 有 21 个共同元素,语料库 A 和 C 有 1 个共同元素;语料库 B 和 C 有 1 个共同元素;语料库 A、B、C 有 22 个共同元素。

3 个语料库的词频可视化分析如图 6 所示。由图 6 可知:3 个语料库相互独立,但有较多相同术语,

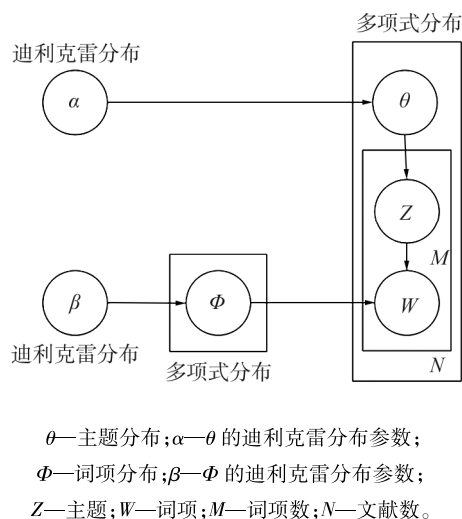


图 4 LDA 主题模型运行原理图

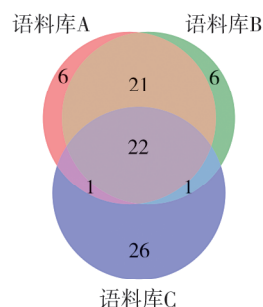


图 5 语料库维恩图

如“大宗商品”“供应链”“风险”“物流”等。说明随文献数的增多,语料库的主题未偏离,3 个语料库均为识别大宗商品供应链风险的有效数据。



图 6 3 个语料库的词频可视化分析

通过文献-术语矩阵分析术语“风险”与其他术语间的相关性,检查相关性较大的术语,筛选得到 87 个与“风险”相关性较高的术语,按相关性从高到低排列,结果如表 7 所示。

表 7 与术语“风险”相关性较高的术语

序号	术语	序号	术语	序号	术语	序号	术语	序号	术语	序号	术语
1	企业	14	国际	27	稳定	40	大宗	53	采购	66	仓储
2	供应链	15	物流	28	业务	41	平台	54	规模	67	库存
3	市场	16	资金	29	政策	42	优势	55	出口	68	运输
4	大宗商品	17	服务	30	控制	43	创新	56	供应商	69	上下游
5	价格	18	成本	31	技术	44	经营	57	资源	70	经济体
6	经济	19	波动	32	合作	45	能源	58	美国	71	原材料
7	中国	20	信息	33	客户	46	冲击	59	信用	72	不确定性
8	贸易	21	产业	34	银行	47	商品	60	战略	73	制造业
9	金融	22	产品	35	货物	48	消费	61	销售	74	质押
10	全球	23	产业链	36	投资	49	核心	62	资产	75	美元
11	管理	24	疫情	37	数据	50	进口	63	粮食	76	现货
12	生产	25	交易	38	环境	51	上涨	64	社会	77	价格上涨
13	需求	26	融资	39	安全	52	监管	65	供给	78	俄乌冲突
										79	农产品
										80	流动性
										81	通胀
										82	汇率
										83	金融机构
										84	美联储
										85	互联网
										86	风险管理
										87	政治

以相关性较高的术语“价格”为例分析对供应链风险管理的影响,价格是大宗商品供应链管理中的重要部分,可从多方面影响供应链的稳定性。大宗商品的价格通常受市场供需、货币政策等多种因素影响,大宗商品企业通常采用衍生工具,如期货合约对冲价格风险,但大宗商品价格极端波动会使套期保值困难,增加套期保值成本,影响企业风险管理效率。

通过 Bi-gram 分析识别某术语与术语“风险”连续出现的频率,如“管理风险”“政策风险”“数据风险”等,按出现频率从高到低排列,结果如表 8 所示。

Bi-gram 网络图能直观展示术语间的关系,快速识别供应链风险的关键术语及术语间关系。Bi-gram 分析网络图如图 7 所示。每个节点代表 1 个术语,节点间的线(边)表示这些术语的共现关系,圆圈表示出现频率较高的节点。图中物联网技术和区块链技术是大宗商品贸易常用的 2 项技术;俄乌冲突和新冠疫情分别反映地缘冲突和公共卫生事件对大宗商品供应链的影响;世界经济、全球经济、中国经济反映大宗商品供应链的全球化特点。

降维后的两个主要维度,用来可视化高维空间中主题间距离的低维表示。通过多维缩放方法确定气泡的位置,在二维平面上展示主题间的距离关系,主题靠得越近,语义越相似。

LDA 主题模型困惑度随主题数的变化趋势如图 8 所示。由图 8 可知:随主题数增多,LDA 主题模型的困惑度整体呈下降趋势;主题数增至 13 后,困惑度的变化趋于平缓并逐渐稳定。将曲线中出现明显拐点主题数 13、14、18、19、20、23、27 分别输入 LDA 主题模型进行试验,通过 pyLDavis 将结果可视化。主题数为 13、14、18、19、23、27 时,pyLDavis 可视化主题效果图中出现明显的象限分布不均匀、主题重合的情况。主题数为 20 时的可视化结果如图 9 所示,主题气泡交叉情况较少且象限分布均匀,主题间相对独立,因此确定最优主题数为 20。

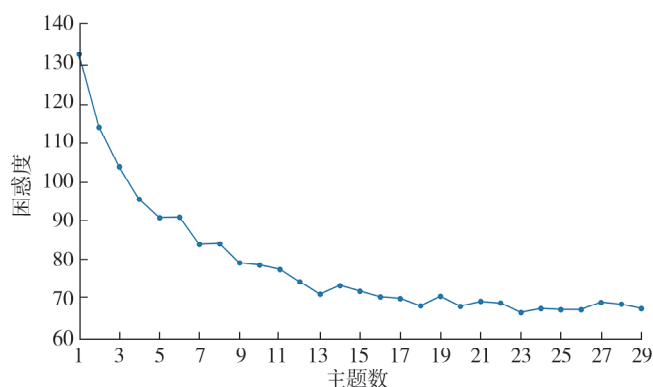


图 8 LDA 主题模型困惑度随主题数的变化趋势

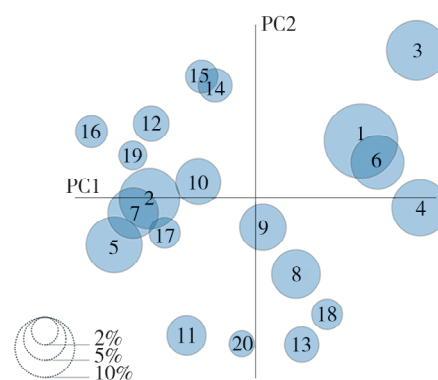


图 9 主题数为 20 时的 pyLDavis 主题聚类可视化分析

通过 LDA 主题模型生成 20 个大宗商品供应链风险主题,每个主题列出 7 个关键词,结果如表 9 所示。

表 9 大宗商品供应链风险主题和关键词

编号	主题名称	关键词
1	市场价格与成本分析	价格、成本、市场、期货、波动、原材料、库存
2	疫情影响与全球经济	疫情、经济、全球、零售商、货币政策、原材料、物流
3	物流优化与农产品供应	物流、农产品、运输、物资、供给、仓储、港口
4	全球冲突与经济影响	全球、中美关系、经济、冲突、俄乌冲突、贸易制裁、地缘政治
5	金融创新与物联网技术	供应链金融、物联网、信息、贸易、区块链、交易市场、数据
6	商品安全与质量监管	指标、商品安全、检验、监管、安全、标准化、追溯
7	价格波动与期货市场	价格波动、期货、市场、全球、经济周期、价格、合约
8	自然灾害与全球经济	地震、日本、海啸、经济、全球、中断、损失
9	粮食安全与产业链稳定	粮食、产业链、安全、石油、市场、政策、波动
10	国际贸易冲突与投资影响	国际贸易、投资、贸易战、汇率、政策、中国、波动
11	金融业务与贸易交易	金融、贸易、交易、银行、业务、供应链金融、市场
12	市场经济与政策稳定性	市场、经济、价格、政策、稳定、波动、供需
13	金融服务与金融诈骗	供应链金融、信用、仓单质押、爆雷、诈骗、应收账款、融资
14	贸易数据分析与信用评估	贸易商、数据、信用、物流、流水、创新、指标
15	全球债务风险与货币政策	债务、加息、美元、美联储、世界、经济周期、汇率
16	贸易业务与国企管理	贸易、客户、国有企业、管理、客户、监管、政府
17	通胀预期与经济政策	经济、通胀、全球、预期、美国、货币政策、金融
18	物流风险与成本管理	运输、延误、损坏、仓储、成本、中断、货物
19	信用风险与合作风险	违约、信用、信息、道德、合作伙伴、上下游、冲突
20	企业经营与管理	战略、数据、技术、决策、操作、管理、企业

LDA 主题模型揭示语料库 B 中文献的主要主题,每个主题从不同角度展示当前大宗商品供应链面临的风险,如,“地震”“海啸”“中断”等关键词从自然灾害角度反映大宗商品供应链面临的风险;“仓单质押”“应收账款”等关键词从金融服务角度反映大宗商品供应链面临的风险;“战略”“决策”“管理”等关键词从企业经营角度反映大宗商品供应链面临的风险。

3.2 大宗商品供应链风险识别结果分类及对比

基于词频分析、相关性分析、N-gram 分析和 LDA 主题建模的结果,将大宗供应链风险分为市场风险、物流风险、金融风险、环境风险、管理风险、合作风险 6 种类型,每种供应链风险类型下又根据关键词细分具体的风险因素,结果如表 10 所示。

表 10 大宗商品供应链风险类型及风险因素识别结果

风险类型	风险因素	相关词频
市场风险	价格波动	价格下降、价格上涨、销售价格、交易价格、价格波动
	汇率波动	汇率波动、货币汇率、货币贬值、外汇市场、波动
	周期性风险	经济衰退、通货膨胀、经济周期、全球化
物流风险	物流成本	供应链管理、燃油价格、运输成本、仓储成本、运费
	仓储风险	库存过剩、库存管理、库存短缺、库存损耗、仓储设施
	运输风险	延误、货损、拥堵、港口积压、苏伊士运河
金融风险	财务风险	现金流量、融资、贷款、坏账、财务风险
	金融业务	欺诈风险、信贷风险、失信、诈骗、重复质押
	核心企业	核心企业资质、核心企业财务状况、核心企业
环境风险	自然灾害	自然灾害、地震、海啸、台风、疫情、中断
	地缘政治	俄乌、贸易制裁、外交、关税、地缘政治
	政策法律	出口限制、税收政策、贸易协议、反垄断
管理风险	质量风险	质量风险、退货、安全事故、批次召回
	技术风险	创新、生产效率、技术、自动化、市场响应
	人员风险	操纵失误、人员素质、罢工、专业、人员风险
	信息风险	信息风险、数据、系统、信息不对称、配对率
	货权风险	货权风险、保险索赔、代理、货损、存货跌价
	战略风险	战略风险、管理、决策、市场定位、投资
合作风险	信用风险	客户资信、逾期账款、违约、信用风险、欺诈
	合作关系	合作关系、信息共享、行业竞争、上下游
	供给风险	供应中断、供给不足、供给风险、原材料
	需求风险	需求预测、客户需求、需求风险、订单减少

传统供应链风险识别方法的识别结果如表 11 所示。对比表 11 结果可知:本文采用的文本挖掘方法与传统风险识别方法的识别结果具有较强的耦合度,识别结果基本包括其他方法识别的供应链风险;文本挖掘方法在基本方向未偏离的基础上具有丰富的拓展性,识别维度更全面,充分体现大宗商品供应链特性,如核心企业风险、货权风险、金融业务风险是其他风险识别结果未识别出的风险。

对文本挖掘识别到的大宗商品供应链类型和因素按层次结构分类,得到的风险类型因素不独立,几个风险因素或风险事件会产生连锁反应,影响部分或整个供应链,供应链的基本信息和组成(行业类型、产品类型、涉及地区等)对发现这些风险因素间的联系至关重要。虽然识别结果未揭示风险因素间的内在联系,但了解大宗商品供应链不同的风险类型及其潜在风险因素至关重要。

表 11 传统供应链风险识别方法及结果

风险识别方法	风险识别结果
德尔非法、流程法 ^[5]	供应风险、生产流程风险、物流配送风险、外部环境风险
文献分析法 ^[8]	生产管理风险、企业管理风险、对接合作风险、技术风险、外部环境风险
业务模式分析法 ^[24]	市场风险、信用风险、外部环境风险、货权风险、人力资源风险、信用风险、操作风险、系统性风险、环境风险
业务模式分析法 ^[25]	生产风险、政治风险、经济风险、运输风险、外部依赖风险、进口集中风险、价格波动风险、金融市场风险
SCOR 模型 ^[26]	自然环境、宏观经济、决策制定、供应商选择、员工流动、资金流动、运输能力、库存技术、信息风险、服务质量
SCOR 模型 ^[27]	人员素质、自然灾害、价格波动、内部环境、质量管理、设备故障、生产事故、运输工具、配送信息等
故障树模型 ^[28]	供应商风险、库存系统风险、物流系统风险、外部环境风险

4 结束语

为识别大宗商品供应链风险,基于文本挖掘方法,提出包括数据采集及语料库构建、数据预处理、风险识别 3 部分的大宗商品供应链风险识别模型框架,通过词频分析、N-gram 分析、相关性分析、累计 TF-H 降维及 LDA 主题建模,识别 20 个大宗商品供应链风险主题,将识别的大宗商品供应链风险分为市场风险、物流风险、金融风险、环境风险、管理风险和合作风险 6 类风险类型,与传统供应链风险识别方法的识别结果对比可知,文本挖掘方法识别维度更全面,识别结果更准确。所建立的大宗商品供应链风险识别模型从大数据视角出发为供应链风险识别提供更科学的工具和方法,使风险管理从传统的经验判断和定性分析转向更科学的数据驱动决策。

本文提出的风险识别模型框架具有可拓展性和可复制性,通过更换数据源,可有效识别其他领域的供应链风险因素。风险识别作为供应链风险管理的第一步,识别结果直接影响风险评估和风险控制的效果和效率,未来可基于风险识别模型识别的风险因素作进一步的风险评估和风险控制研究。

参考文献:

- [1] 史乐蒙,宋华.大宗商品供应链必须转型[N].期货日报,2023-08-28(7).
- [2] 尹志.国际大宗商品价格波动及对中国的启示[J].价格月刊,2023(11):16-22.
- [3] 方先明,高元.地缘冲突风险、大宗商品金融化与农产品期货价格波动[J].经济问题,2023(6):57-67.
- [4] 宋华,韩思齐,刘文诣.数字化金融科技平台赋能的供应链金融模式:基于信息处理视角的双案例研究[J].管理评论,2024,36(1):264-275.
- [5] 张敏,杨阁,史一鸣.直播电商模式下农产品直播供应链风险识别与测度研究[J].物流科技,2024,47(17):121-126.
- [6] 李国亮.基于 SCOR 模型的 HZDH 公司供应链风险识别与控制研究[D].桂林:桂林理工大学,2023.
- [7] CAO J, YE X H, QI Y, et al. An integrative decision-making model for the operation of sustainable supply chain in China [J]. Energy Procedia, 2011, 5: 1497-1501.
- [8] 谢乐.基于贝叶斯网络的农超对接供应链风险预警模型研究[D].南昌:江西财经大学,2020.
- [9] AQLAN F, LAM S S. A fuzzy-based integrated framework for supply chain risk assessment[J]. International Journal of Production Economics, 2015, 161: 54-63.
- [10] HASSAN A P. Enhancing supply chain risk management by applying machine learning to identify risks[C]//Proceedings of International Conference on Business Information Systems. Seville, Spain: Springer Nature Switzerland AG, 2019: 191-205.
- [11] 牛莉霞,赵蕊.大数据时代煤矿安全风险治理模式研究[J].煤矿安全,2022,53(7):241-245.
- [12] ZHAO L T, GUO S Q, WANG Y. Oil market risk factor identification based on text mining technology[J]. Energy Procedia, 2019, 158: 3589-3595.
- [13] 刘心男,王家慧,纪颖波.建设工程质量检测机构运行风险管理研究[J].建筑经济,2023,44(11):89-94.

- [14] CHU C Y, PARK K, KREMER G E. A global supply chain risk management framework: an application of text-mining to identify region-specific supply chain risks[J]. *Advanced Engineering Informatics*, 2020, 45:101053.
- [15] 犹梦洁. 基于文本挖掘的煤矿安全风险识别与评价研究[D]. 徐州:中国矿业大学,2022.
- [16] 郑福,何璐彤,洪灵辉,等. 基于文本挖掘的电商平台茶类商品消费者满意度研究:京东商城苦丁茶在线评论为例[J]. *商展经济*,2024(4):121-125.
- [17] 郑秀婷. 文本隐含气候风险因子定价[D]. 武汉:华中科技大学,2021.
- [18] 林晓鹏. 基于 N-gram 算法的网络安全风险检测系统设计[J]. *信息与电脑(理论版)*,2023,35(15):215-217.
- [19] 郭家璇. 基于 TF-IDF 算法的货运平台车货供需匹配研究[D]. 北京:北京交通大学,2022.
- [20] XU N, MA L, LIU Q, et al. An improved text mining approach to extract safety risk factors from construction accident reports[J]. *Safety Science*, 2021, 138:105216.
- [21] 于胜博. 考虑消费者体验感的生鲜农产品配送系统优化研究[D]. 南昌:江西财经大学,2021.
- [22] 陈芳,陈茜,徐碧晨. 基于文本挖掘的管制运行风险主题分析[J]. *中国安全生产科学技术*,2020,16(11):47-52.
- [23] 陈芳,温抗抗,张亚博,等. 民航管制安全风险主题时空分布规律研究[J]. *安全与环境学报*,2024,24(2):587-595.
- [24] 曾尧. 大宗商品供应链管理企业风险管理及其评价研究[D]. 北京:北京交通大学,2021.
- [25] 丁存振,徐宣国. 国际粮食供应链安全风险与应对研究[J]. *经济学家*,2022(6):109-118.
- [26] 许保光,谢传胜. 煤炭企业供应链风险识别与控制[J]. *中国储运*,2024(1):88.
- [27] 胡梦婷. 基于贝叶斯网络的汽车供应链风险识别与评估[J]. *物流技术*,2023,42(8):137-142.
- [28] 徐园林. 基于贝叶斯网络的装配式建筑供应链风险分析[J]. *项目管理技术*,2024,22(5):125-132.

Risk identification of bulk commodity supply chains

ZHAO Xin, QI Danyang, LIU Huaqiong*

School of Transportation and Logistics Engineering, Shandong Jiaotong University, Jinan 250357, China

Abstract: To address issues in traditional risk identification methods for commodity supply chains, such as incomplete perspectives and low accuracy of identification results, a text mining approach is employed to establish a bulk commodity supply chain risk identification model framework, which includes data collection, corpus construction, data preprocessing, and risk identification. Research papers related to bulk commodity supply chain management are collected from China National Knowledge Infrastructure (CNKI) and Wanfang Data Knowledge Service Platform. Three corpora with different numbers of texts are constructed. These corpora undergo word frequency analysis, N-gram analysis, correlation analysis, term frequency-information entropy (TF-H) dimensionality reduction, and latent Dirichlet allocation (LDA) topic modeling. The results of the risk identification are compared with those from traditional supply chain risk identification methods to validate the effectiveness of the proposed approach. The results show that the LDA topic model generates 20 bulk commodity supply chain risk topics, each reflecting the risks faced by the current bulk commodity supply chain from different perspectives. The identified risks are categorized into six types: market risk, logistics risk, financial risk, environmental risk, management risk, and cooperation risk. The text mining approach demonstrates a strong correlation with traditional risk identification methods, while offering a more comprehensive identification dimension and more accurate results. Text mining technology can comprehensively and accurately identify supply chain risk factors and provide theoretical support for bulk commodity supply chain risk identification.

Keywords: bulk commodity; supply chain management; risk identification; text mining; LDA

(责任编辑:赵玉真)