

大数据技术对市场监管的启发

——数据驱动的质量风险评估

郑立伟^{1,2,3}, 刘中华^{2,3}, 倪超^{2,3}

(1.清华大学工业工程系,北京100084;2.中国航空综合技术研究所,北京100028;

3.国家市场监督管理总局质量基础设施效能研究重点实验室,北京100028)

摘要:质量信息难以获取一直是政府质量治理面临的重大挑战。互联网及大数据技术的发展使多元异构数据的采集、处理及动态分析成为可能。聚焦基于大数据的质量风险评估,提出了数据驱动的质量风险评估思路,系统性地识别供应风险、经营风险、准入风险、诚信风险和技术风险五种不同模式的风险来源,通过散落在互联网中的数据构建了不同模式风险来源的表征指标,最终设计了质量风险评估系统框架,并以产品缺陷风险预警为例验证了框架的实用性和有效性。研究结果为大数据技术在质量风险评估中的应用提供了有益的启示。

关键词:质量风险;数据驱动;风险评估;质量治理

中图分类号:F253.3 **文献标志码:**A **文章编号:**1002—980X(2021)08—0033—08

一、引言

由于市场监管存在漏洞、消费者权益保护不足、未成熟新技术“非适度”应用,以及质量违法与假冒伪劣制售行为扩张等因素的不利影响,在生产与消费环节存在着大量的质量风险。这些风险早期以偶发的形态出现,但如果缺乏有效的识别、监测和治理,这些风险中的一部分就会沿着产业链或消费链传递、聚集和扩张,逐步演化为区域市场中的巨大风险,最终涌现为突然爆发的重大质量安全事件。因此,对质量风险进行识别和管理,已成为世界各国市场监管及政府质量治理的主要任务,也是各国的难以有效处理的共性问题。欧盟、美国都建立了缺陷产品的快速通报系统。我国市场监管部门及行业管理部门也开展了产品质量风险监测。其核心是通过抽样检验,监测产品的质量安全状况,并对质量风险信息予以通报,防止大范围发生质量安全事件。质量监管的难点在于市场信息的不对称性。政府在质量监管等环节积累了大量的数据,如监督检查、行政执法等数据;互联网中存在大量“散落的”质量信息,例如消费者的投诉、质量新闻等。一方面,大数据应用于质量风险评估,利用多维数据的采集、分析、传播的先天便利性极大程度的弥补了市场信息的不对称性;另一方面,通过大数据技术整合互联网中的质量信息可以快速地捕捉质量信息,以提高质量风险评估的及时性和动态性。与此同时,应用大数据技术推动政府管理和社会治理模式进步是国务院在《促进大数据发展行动纲要》中提出的要求,也是顺应时代技术趋势的必然选择。质量是一个多主体参与的系统性结果,有效的预警工具和方法对质量风险的管理水平尤为重要。本文探讨应用大数据技术开展数据驱动的质量风险评估,提高我国质量风险管理的动态性、及时性和有效性。

二、文献回顾

“风险”一词源于客观危险,现代意义上的风险被赋予了更加深刻的含义。由于学者对风险研究角度的差异,风险的内涵至今还没形成统一的定义,可归纳出三种代表性观点。①风险是事件未来可能结果发生的不确定性。Mowbray(1995)和Williams(1985)将风险定义为特定条件下结果的不确定性。②风险是损失发生的不确定性。Rosenb(1972)和Crane(1984)将风险定义为损失的不确定性。结果不确定性不一定导致损

收稿日期:2020—12—08

基金项目:北京市科学技术委员会城市运行安全保障项目“基于大数据的产品质量安全风险预警技术与平台建设”(Z191100001419016)

作者简介:郑立伟,清华大学博士研究生,中国航空综合技术研究所研究员,研究方向:大数据、质量管理、质量统计与评价;刘中华,硕士,中国航空综合技术研究所助理工程师,研究方向:数据挖掘、质量风险预警;倪超,博士,中国航空综合技术研究所高级工程师,研究方向:宏观质量评价。

失的发生,进一步将风险的内涵精准化。③风险是指损失的大小和发生的可能性。阎春宁(2002)将风险定义为:“风险是在特定的客观情况和时期内,预期结果与实际结果之间的差距及变动程度的概率分布”。

风险在金融行业中应用起步较早,较为成熟,但学术界对质量风险问题的认识还不够,相关的研究还处在起步阶段。对质量风险的研究也更多集中于对具体产品在生产、消费及供应链环节的质量风险识别与评估方面。Porthin et al(2018)提出了一种启发式算法用于对纸张制造的质量风险进行评估,算法通过对专家风险的评估结果进行结构化分析,实现了最优风险判定和风险控制效果;Gereffi和Lee(2009)针对全球性的食品安全事件,建立了基于全球价值链的食品标准和食品质量风险分析模型,发现价值链结构、食品标准、市场导向性质等因素密切影响了食品质量风险。从国内看,相关研究主要集中在食品类质量安全评估领域。孙君茂(2007)和杨艳涛(2009)研究了食品质量安全风险评估,着眼于食品生产环节,以某一区域的食物质量安全水平(区域食物质量安全度)为因变量,以所有关于该区域食物质量安全影响因素为自变量的函数,并明晰了影响因素的作用机理;安珺(2012)运用层次分析法,构建了乳品质量安全预警指标体系并开发出具有数据快速查询、信息共享和快速预警功能的乳品质量安全预警系统。金王平(2013)和Wang和Yue(2017)用无限射频识别(RFID)传感器技术和物联网技术,通过针对乳制品的供应链和生产链的全流程监测,实现针对企业乳制品质量风险的监控和预警。

大数据技术在很多领域都得到了应用,但在质量治理中,大数据技术的研究刚刚兴起,大部分停留在理论层次,应用尚不成体系。许多学者从理论上构建了大数据技术应用于质量评价、质量分析的路径和模式,对如何进行风险预警研究较少,一些学者将大数据应用于服务质量评价、产品伤害监测和产品召回及如何基于大数据的产品质量监管新模式(何云福,2016);商允伟和梁娜(2016)提出了基于大数据的产品质量分析路径。倪亚晖(2018)通过定性与定量相结合的方法界定了质量风险信息,并确定了质量风险信息的主要来源及采集方式,但未提出如何对采集的质量风险信息进行应用。庞军朋(2018)分析了欧盟和美国的产品质量风险评估方法,对建立产品质量风险分析评估方法进行了初步研究,仅展望了产品质量风险评估,并未进一步深入研究。郭先超等(2016)通过大数据MapReduce技术实现了对产品质量风险的评估,但只考虑抽检数据,并未充分利用互联网上的“质量”信息。蔡佳苗等(2014)等从信息系统的角度研究如何整合产品质量信息系统来解决智能分析和预测问题,但未对评估的理论和方法进行深入研究。少数学者在产品、环境、教育等领域开展了运用大数据进行质量治理的应用研究。钟攀等(2015)基于四川省农产品质量安全风险监测平台建设经验及大数据分析模型,提出了风险监测的模式创新路径和政策建议;王锋等(2017)通过研究高等教育质量监测大数据的分布式文件存储技术、并行处理技术和数据挖掘技术,为实施大数据驱动的高等教育质量监测评估奠定了关键技术基础。肖湘雄(2015)提出利用大数据提高农业标准化程度,提升大数据技术水平,完善农产品质量安全治理法律法规等建议。

我国对产品质量安全的监管局限于监督检查等监管手段,对产品的缺陷风险评估通过检验检测技术对实体产品进行买样测试,缺乏时效性;与此同时忽略了互联网上较为及时的质量信息。学者对产品质量风险的研究起步较晚,应用研究不成体系。大数据技术在质量监管的应用中,更多的关注行业性的质量分析,对质量风险的预警较为缺乏。在此背景下,本文聚集基于大数据的产品的质量风险评估,为大数据技术在质量风险评估的应用提供有益的启示。

三、数据驱动的质量风险评估

(一)数据驱动的质量风险评估思路

历史经验证明,经济环境对质量风险有系统性的作用。经济环境通过作用于市场环境进而影响质量风险。20世纪90年代初期,我国由供给制约型经济全面转向需求制约型经济,在经过一轮过热的建设高潮后,经济增长速度和通货膨胀迅速攀升,1997年爆发的亚洲金融危机,使我国经济在“软着陆”之后,感到了需求不足的巨大压力。90年代末,我国首次出现假冒伪劣盛行的状况。李英东和俞炜华(2008)研究表明,当前我国从经济驱动因素、产业结构、人均国内生产总值等方面与美国20世纪初非常相似。美国正是在这一时期也大范围爆发质量问题。厄普顿·辛克莱尔的《丛林》(The Jungle)一书进行了形象的描述。

社会需求层次、市场竞争程度、科学技术水平和管理精细程度从根本上决定产品质量的高低。经济发展、科学技术、社会管理等薄弱环节也必然以质量为载体表现出来。质量风险具有综合性和隐蔽性的特征。

综合性是指影响产品质量的因素复杂。例如,“三聚氰胺”事件既有行业过度竞争产品价格与成本背离的影响,也有奶源加工与乳制品加工产业发展的不匹配,标准和检测存在漏洞等原因。隐蔽性是指诱因和结果具有非直接的关系。例如,丰田召回事件表面是质量缺陷,实际是在产能迅速扩张、压缩成本的同时,质量管理水平下降,放松了对供应商的控制。

目前的质量风险监测方法从本质上还是从微观分析风险,并没有实现利用宏观数据进行预测、预警的目的。现今,一些经济学者已开始从质量入手剖析我国经济发展的弊端。但由于长期以来,质量研究人员对经济的不敏感,还缺少从经济和市场角度,研究经济发展对产品质量的作用机制。

从治理工具看,大数据技术的发展为质量风险的评价与预警提供了新的技术方法与实现路径。由于质量风险从初步出现到最终爆发之间往往存在一段酝酿时期,期间显示风险形成和发展的局部信息会在互联网上以各种形式表现出来,例如,消费者的抱怨与不满,消费者非正式的投诉,消费者因产品缺陷伤害而向医疗机构、法律服务机构及政府相关部门所做的咨询、求助和举报,各级政府质量及市场监管部门在产品质量监督抽查中发现的不合格,以及有关检验、检测与认证机构在技术服务中记录的不符合项等。在过去,这些质量风险信息是以碎片的形式散布在互联网及相关机构的数据库中,无法得到充分利用。例如,2008年的三聚氰胺事件,早在2004年就已在阜阳的农村造成恶果,并被安徽省政府发现和处理。但在2004—2008年期间,尽管相关风险信息仍被不断暴露,但国家有关质量监管部门(包括质检、工商、农业、卫生及工信部门)未能及时识别这一风险从微观形态向宏观形态演化的趋势,坐视零散性、局部性、偶然性的质量安全问题最终发展成为区域性、产业性及系统性的重大质量安全问题,并在2008年彻底爆发;又如,2017年的西安地铁电缆事件,西安市质量监管部门早在2015年就收到过奥凯电缆问题的投诉,并进行了不合格结果公示。同期,互联网上也出现了奥凯电缆存在质量问题的部分“杂音”,但地铁主管部门却忽视了此类信息,成为助推相关事件不断向坏演化的内在原因。这些真实案例表明,由于宏观质量风险的“预兆信息”碎片化、散落化特征,对其的识别、评价与预警研究更适合采用大数据技术手段,可以以仿真建模和量化评价的方式,通过“多源多态多制数据收集和处理→质量风险特征分析→质量风险识别、评价与预警→治理模式与策略研究”这一链条,对大量孤立的微观质量风险相关信息源进行汇聚、整合和深度挖掘,从而评估、预测出宏观质量风险的演化趋势及爆发系统性质量安全突发事件的可能性,支撑政府和消费者进行有效的风险规避、化解和治理,数据驱动的质量风险评估思路如图1所示。

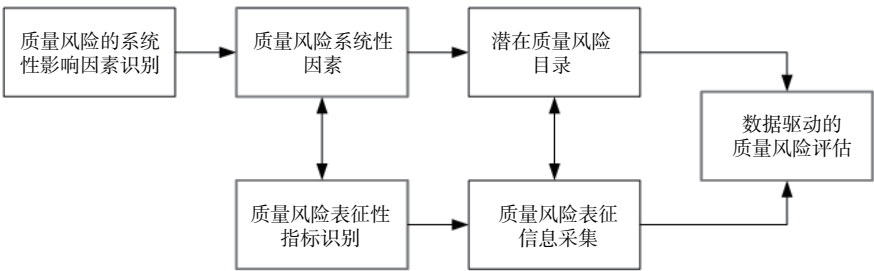


图1 数据驱动的质量风险评估思路

- 数据驱动的质量风险评估的关键技术主要有3个方面：
- (1)质量风险的影响因素分析;系统性地分析影响质量风险的因素,质量风险影响因素分析是数据驱动的产品质量风险评估的基础。
 - (2)构建质量风险因素的表征指标;质量风险因素的表征指标是通过散落在互联网中的信息搭建数据与风险因素两者关系的桥梁,风险因素表征指标是风险因素的“触角”,通过“触角”的变化感知风险的走向。
 - (3)建立基于互联网多元异构数据的质量风险评估系统;通过大数据平台技术,监测可识别表征指标的稳定数据源,分析预警质量风险,对质量风险进行推送与报警。

(二)质量风险的影响因素及表征指标

1. 质量风险的系统性影响因素

产品质量形成的主体是生产企业,在利益、信息、供应链、经营模式等因素的作用下,质量风险不仅仅局限于生产活动。从生产企业的视角来看,产品全生命周期所涉及各个环节为“生产主体建立→新产品开发→产业化→批量生产→运输→销售”,多主体参与其中主要涉及生产企业、监管机构及第三方机构(标准、计量、

认证机构等)。监管机构主要对企业和产品的准入监管及流通领域的诚信监管;生产主体从供应链角度来看,涉及原材料采购、加工、运输等环节,并主要负责产品的经营过程。第三方机构提供标准、计量和认证等活动内容,给监管和消费者传递质量信号。从产品形成全生命周期来看,质量风险可分为五大类,这些风险的来源是影响质量风险的系统性因素。

(1)供给风险。供给风险(supplying risk),上游产业链造成的质量风险。原材料的质量、价格沿着产业链向下游产业传递。原材料/元器件质量影响最终产品质量,如钢板质量影响汽车质量。“双汇”火腿肠中检测出“瘦肉精”,按照产业链分析,除了火腿肠外,熟肉制品和肉类罐头都可能存在“瘦肉精”。原材料非常稀缺或昂贵的产品面临较大的供给风险,如假紫砂。猪肉价格上涨,食品企业迫于生存压力和价格管制,利润有所下降,将直接加剧屠宰业、肉制品加工业、肉类罐头制造业和餐饮业的质量安全隐患。

(2)经营风险。经营风险(running risk),市场环境造成的质量风险。供需结构影响市场环境。产能过剩将导致市场竞争异常激烈,企业大面积亏损,出口不计成本,国际贸易摩擦频繁发生等问题。生产企业为了降低生产成本可能选择低价原材料、简化工艺等,产品质量将下降。如果退出成本过高,行业将在长时间内维持恶性竞争。相反,产品供给缺口较大,销售价格就会大幅上涨。如果行业准入门槛低,将催生假冒伪劣市场。同时,产能、市场需求的迅速扩张或萎缩都导致供需结构的急速变化,在短期内造成产品质量的波动。

价格是市场最灵敏的信号。在健康的市场环境中,价格信号与价值信号基本是等价的。原材料价格、生产成本、物流成本的波动都会体现在产品销售价格的变化上。当市场错误地传递价值信号,对于价格的敏感程度远远强于对于产品质量的敏感程度,产品销售价格和成本相背离,整个产业就面临经营风险。销售价格和生产成本的剧烈波动都会导致产品质量风险增加,例如,2020年,口罩价格急剧飙升,口罩的掺假、造假事件层出不穷;2016年,北京抽检广东生产小家电产品八成不合格,究其原因小家电企业压缩成本,低价竞售,没有质量保证的产品只能降价竞争,从而形成恶性循环。

(3)准入风险。准入风险(admittance risk),准入条件造成的质量风险。准入要求是为了保护使用者的人身安全和利益,由政府部门、行业组织等设定的进入某区域市场的门槛,即产品应强制达到的水平,也是调节产能结构的主要手段之一。准入要求依据已知和未知的产品风险制定,包括强制性标准、生产许可、产品认证、体系认证和人员认证等。准入要求高则质量风险小,准入要求低则质量风险大。由于技术和社会发展水平的差异,各个国家、区域对同一产品的准入要求存在差异。一般来说,经济相对落后国家为了保护本国产业发展的需要,以及受到消费水平的限制,准入要求往往比发达国家低。我国一直是世界贸易组织/技术性贸易壁垒协定(WTO-TBT)通报的重灾区,这虽然与我国出口数量较多有关,但也反映了我国对于部分产品的质量要求与发达国家还有差距。

(4)诚信风险。诚信风险(credit risk),企业诚信水平造成的质量风险。企业的诚信和道德是质量的基础也是最后防线。企业注重诚信会抵消各方面的不利因素,降低风险。企业越缺乏诚信,质量安全越脆弱。诚信水平与违法、违规成本相关,违法成本越高,企业越讲诚信,相反,成本太低将纵容企业的不诚信行为。违法成本由法律法规惩罚力度、政府监管力度、社会监督力度和企业发展阶段决定。企业发展越成熟,品牌影响力越大,因失信造成的损失就越大,也就更约束企业维持较高的信誉水平。我国社会缺乏诚信已经受到全世界的诟病,一些企业丧失了基本的行业道德,这与我国处罚较轻,企业的规模小、品牌影响力低有很大关系。

(5)技术风险。技术风险(technology risk),产品技术问题造成的质量风险。技术风险来源于产品的技术成熟度和质量的技术复杂性。技术越成熟,不可预知的产品质量问题越少。在产业发展初期,产品的各种质量问题没有充分暴露,质量风险较大,如现阶段新能源汽车、高速列车。质量的技术复杂性由产品的质量特性决定,产品质量特性可分为物理特性、物质含量、电磁特性和理化特性等,常见产品的质量特性见表1。一般来说,物质含量和理化特性的技术复杂性较低,物理特性和电磁特性技术复杂度较高。技术复杂度越高,产品质量风险越大。

表1 常见产品的质量特性

产品质量特性	物理特性	物质含量	电磁特性	理化特性
常见指标	机械性能	甲醛、添加剂	辐射	色牢度、色差
典型产品	玩具、机床	食品、药品、服装	手机	服装

2. 质量风险表征性指标识别

根据质量风险的系统性影响因素,可形成评估潜在质量风险的指标,根据质量风险暴露的情况可形成质量风险的表征性指标。一方面,建立潜在的质量风险评估指标,可以通过稳定的监测评估指标以期识别潜在

的质量风险问题,潜在质量风险的评估指标见表2。对于供给风险主要监测原材料的价格和质量,价格是供需最敏感的信号;当原材料价格的波动异常反映原材料供求关系异常,此时原材料质量问题向产品质量传导的风险较大。产能利用率是衡量产品供需结构和经营环境的理想指标。美联储认为,如果产能利用率达到85%,就可以认为实现了产能的充分利用。超过90%,就可以认为产能不足,有可能引起通货膨胀。如果明显低于79%~83%区间,则说明可能存在产能过剩,即开工不足。成本费用率可以有效的衡量企业收益,进而反映企业的经营状况。当企业成本费用利润率远高于社会平均利润率,就会吸引其他资金进入,假冒伪劣的风险也较大。准入风险主要针对企业的准入和产品的准入,从我国与发达国家准入要求的差距,以及准入监管措施的有效性两个角度识别准入风险。对于诚信风险主要从法律法规的惩处力度及违法的意愿两个方面;监管力度可以有效的遏制诚信风险,比如对于假冒伪劣的严重处罚。小企业是影响整体产品质量的重要因素,与大中型企业相比,其诚信水平、产品质量水平保障能力不足,小企业的产值比重可以反映诚信风险。技术风险与产品联系紧密,对于具体产品的技术风险可以采用技术成熟度、产品质量特性的复杂度等指标进行评估;另一方面,社交网络及互联网平台的广泛应用,消费者或使用者的质量信息在网络传播,使质量信息比以往更加透明。通过消费者、国内外召回、产品伤害、行政执法等偶发性表征指标可以快速的识别偶发的质量风险,质量风险的偶发性表征指标见表3。

表2 潜在质量风险的评估指标

类型	指标示例
供给风险	原材料价格波动
	原材料质量问题
经营风险	成本费用利润率
	产能利用率
准入风险	标准差距
	检测技术差距
	认证差距
诚信风险	监管力度不足
	小微企业产值比重
技术风险	技术成熟度低/产品上市时间

表3 质量风险的偶发性表征指标

类型	来源
消费者投诉	12345、12315、12365等消费者投诉信息
商品不良评价	网购平台及专业产品网站(如,汽车之家、大众点评、旅游评价等)的产品评价和用户反馈
国外召回记录	国外发布的召回信息
产品伤害记录	国内医院记录的产品伤害数据、平台的产品伤害就诊记录
行政执法记录	市场监管部门的行政执法记录
质量舆情信息	社交平台、新闻门户网站的舆情信息

(三)面向数据驱动的质量风险评估系统框架

数据驱动的质量风险评估涉及的数据来源广、数据类型多样,需要进行数据治理以达到整合利用的目的。同时,要实现数据实时采集,对质量风险进行动态评估,建立质量风险评估系统框架,如图2所示。

1. 数据资源层

通过供给风险、经营风险、准入风险、诚信风险和技术风险等系统性风险影响因素及梳理的风险因素表征指标,梳理政府市场监管、国内外产品召回通报、消费者投诉等渠道的数据,确定稳定的数据来源渠道,并明确数据维度。

2. 数据采集层

以Scrapy为核心构建分布式数据采集服务,通过Redis通信各数据采集服务主机以提升数据采集服务效率。针对数据源网站分别在采集服务平台中添加相应数据采集过程,实现对数据的自动采集与监控。面对不同数据的接入需求,采用针对性接入方式及工具:①针对实时数据采用分布式消息队列Kafka进行接入;②针对同步关系型数据库的数据,采用Sqoop全量或定时增量抽取;③针对日志文件,采用Flume聚汇到Hadoop分布式文件系统(HDFS)。

3. 数据共享平台

(1)基础平台。通过Hadoop、Hive、Hbase、Mysql、Mongdb等搭建基础存储平台,实现原始数据和增量数据的存储。

(2)数据治理。建立主数据、元数据管理机制,数据仓库的数据链路加工机制;数据共享中心:集成数据产品中公共、通用的业务,封装数据服务;数据质量管理:在使用过程中发现质量错误,追踪溯源,修改整条链路上的数据。

(3)数据集市。以质量风险预警为主题,建立以产品、企业及区域为核心的数据集市,并通过历史数据和新风险词发现机制建立和持续扩种风险词库。

4. 基于多元异构数据的质量风险预警

融合多源异构的质量风险数据,识别风险特征并监测质量风险的表征性指标,依据大数据平台采集的数据进行自动风险分析,当风险突破阈值时,推送风险预警信息。

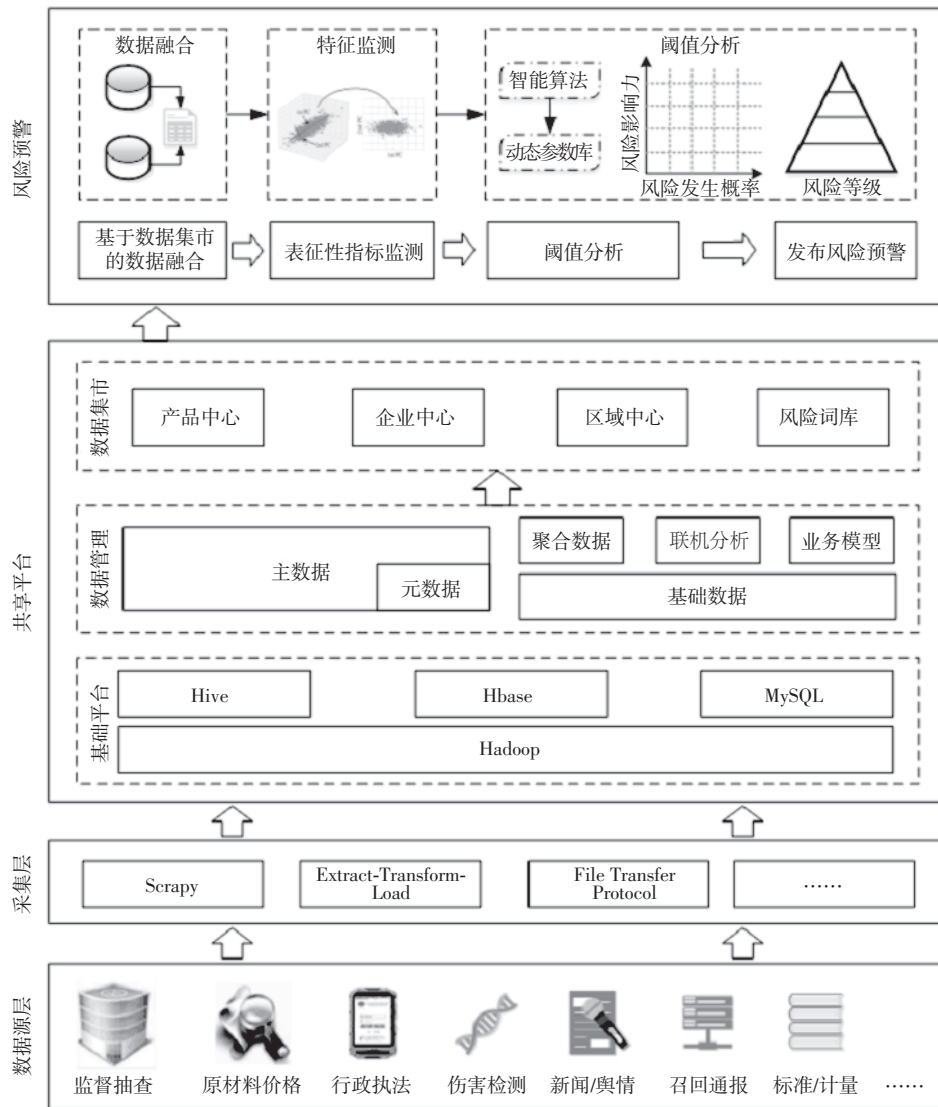


图2 质量风险评估系统框架

四、实证验证

结合 2019 年度国家监督抽查、各省份抽检数据、美国消费品安全委员会 (Consumer Product Safety Committee, CPSC) 召回数据及欧盟非食品类消费产品的快速警报系统 (The Rapid Alert System for Non-food Consumer Products, RAPEX) 通报数据,预警产品缺陷风险,从而验证质量风险评估系统框架的有效性和实用性。

(一) 数据源梳理及数据采集

从政府市场监管、进出口质量等维度出发,寻找国家监督抽查(例如中国质量新闻网)、全国各省份监督抽查(各省市市场监管局网站)、CPSC 产品质量通报、RAPEX 通报等渠道,建立识别缺陷产品的稳定数据源网站。针对不同网站特性及网站数据特征,确定需要采集数据,如监督抽查数据、召回数据等,并明确数据的维度,如产品名称、产品类别、产品生产日期等。通过互联网采用 Scrapy 分布式爬虫技术编写爬虫程序,分别对以上数据进行采集,采集 2019 年全年数据,共收集到 2019 年国抽不合格数据 2787 条、2019 年省抽不合

格数据 7516 条、CPSC 召回数据 754 条和 RAPEX 通报数据 4960 条。

(二)数据接入共享平台

针对 CPSC 通报数据和 RAPEX 数据通过分布式消息队列 Kafka 接入；国抽和省抽按季度进行采集的数据通过 Sqoop 全量或定时增量抽取同步关系型数据库。制定数据治理规则，统一名称及编码，通过数据清洗、数据转化、数据提取、数据计算的步骤将分散、零乱、标准不统一的数据整合到一起，实现多源数据的互联互通；针对数据的结构差异分别采用相应的存储方式，其中，结构化数据存储于 Hive 数据仓库，有一定格式关系的半结构化数据存储于 HBase 列族数据库，非结构化数据直接存储于 HDFS 分布式文件系统中。数据存储后，进行主数据管理，建立产品主数据，统一产品信息编码，执行从创建、变更、废弃的全生命周期管理流程，对已采集的产品名称数据进行清洗，包括识别并修改错误的数据信息、规范数据格式、删除无效数据，规定统一的产品名称，生成统一编码，从而形成初期的产品主数据；在产品主数据初步建立以后，持续进行主数据的收集分析，进行数据清洗和校验，祛除冗余，逐步提高产品主数据的质量，从而解决同一个产品多个名称、数据前后不一致、数据标准不统一等问题。以产品主数据做为唯一标识，整合存储的所有数据，最终形成以产品为中心的质量安全数据库。

(三)产品缺陷识别预警

通过构建的产品主数据对 5 类产品涉及的产品（儿童玩具、家具、家用电器、建筑材料和数码电子）进一步关联匹配和筛选，共提取数据 1780 条数据，将提取的数据按照产品名称聚合，共涉及 140 个产品。美国 CPSC 召回数据中包含产品召回的名称、描述、危险、事故等字段。其中，危险、事故及风险描述包含了大量的文本信息。对获得的数据进行处理使其规范化，成为能够供文本挖掘处理的文本格式。将所收集的数据有关对风险描述部分转化为可识别格式，利用分词技术及关键词识别技术对文本数据进行清洗降噪，剔除不合规范和一致性错误的信息，得到满足分析使用的聚合的产品信息表，见表 4（部分实例）。

表 4 缺陷产品识别信息聚合表（部分实例）

产品名称	国抽不合格次数	省抽不合格次数	CPSC 召回次数	Rapex 通报次数	不合格原因	伤害事故	召回原因	Rapex 所述原因
桌面玩具	1	0	0	1		截至目前，已收到 2 起事故报告	圆形塑料珠可能会破裂，造成窒息的危险	
...	...	...	...	...	...	...	...	...
儿童自行车	13	1	3	6	把立管，链罩，链罩等不合格	截至目前，Pacific Cycle 已接到 3 起塑料盖松驰的相关事故报告	自行车在使用过程中曲柄可能会弯曲或断裂，构成跌倒的危险	可能造成受伤的风险

处理后的数据参考专家经验建立的缺陷产品规则，将缺陷产品分为 3 个等级（高风险、中等风险、低风险），通过不合格原因、召回通报原因等文字描述，标注每个产品的缺陷等级。按照 80% 和 20% 的分配比例切分成训练数据和测试数据。通过建立支持向量机（support vector machines, SVM）分类器，开始对数据进行交叉验证的训练调优，形成了缺陷产品的特征向量，并将训练后的 SVM 分类器在测试数据验证，训练节点的准确度达 96%，测试阶段的准确度达 67%。基于分析的 2019 年数据，通过事故经验数据及专家经验比对，高风险产品确实多发产品质量事故，如桌面玩具、摇铃玩具等产品。以产品缺陷风险可以有效的证明质量风险评估系统框架的有效性和实用性。

五、总结

本文提出了“产品影响因素识别”→“影响因素表征”→“表征数据采集”→“风险评估”的数据驱动的质量风险评估思路。从产品全生命周期参与主体角度出发，梳理了质量风险的系统性影响因素，如供给风险、经营风险、准入风险、诚信风险和技术风险等不同质量风险来源，并建立了质量风险偶发性表征指标，最终基于 Hadoop 大数据平台设计了数据资源层、数据采集层、共享平台层及风险预警层的质量风险评估系统架构，最终以产品缺陷风险为例验证了所提出系统架构的有效性和实用性。本文通过产品缺陷风险一个风险实证跑通了评估系统架构，足以验证了评估系统架构的有效性和实用性；但是对于质量风险预警来说，后续可从所提出的系统性及偶发性风险继续深入预警质量风险。

参考文献

[1] 安珺，2012. 基于层次分析法的乳品质量安全预警系统研究[D]. 沈阳：东北农业大学。

- [2] 蔡佳苗, 陈学章, 巫俊宏, 2014. 大数据在产品质量风险信息管理中的应用研究[J]. 标准科学(1): 82-84.
- [3] 郭先超, 林宗缪, 姚文勇, 2016. 大数据环境下产品质量安全风险评估研究[J]. 信息技术(6): 76-79.
- [4] 何云福, 2016. 大数据视野中的产品质量监管[J]. 上海质量(1): 30-32.
- [5] 金王平, 2013. 原料奶质量安全监控和预警系统研究和开发[D]. 杭州: 浙江大学.
- [6] 李英东, 俞伟华, 2008. 近年来我国经济增长形势与19世纪末20世纪初期美国经济发展特征的比较研究[J]. 学术论坛(3): 139-142.
- [7] 倪亚晖, 2018. 大数据分析在质检工作中的应用——基于大数据的产品质量风险信息监测与应用[J]. 电脑知识与技术, 14(32): 254-255.
- [8] 庞军朋, 2018. 大数据技术在产品质量风险评估中的应用初探[J]. 现代商业(21): 30-31.
- [9] 商允伟, 梁娜, 2016. 产品质量状况分析的“大数据路径”[J]. 信息化建设(7): 54-57.
- [10] 孙君茂, 2007. 区域食物质量安全风险评估研究[D]. 北京: 中国农业科学院.
- [11] 王锋, 王翔宇, 秦文臻, 2017. 大数据驱动的高等教育质量监测评估关键技术研究[J]. 黑龙江高教研究(6): 80-83.
- [12] 肖湘雄, 2015. 大数据: 农产品质量安全治理的机遇、挑战及对策[J]. 中国行政管理(11): 25-29.
- [13] 阎春宁, 2002. 风险管理学[M]. 上海: 上海大学出版社: 4-6.
- [14] 杨艳涛, 2009. 加工农产品质量安全预警与实证研究[D]. 北京: 中国农业科学院.
- [15] 钟攀, 葛荣, 杨文, 2015. 农产品质量安全风险监测大数据分析策略与应用研究——以四川省质量安全风险监测为例[J]. 农产品质量与安全(4): 8-12.
- [16] GEREFFI G, LEE J, 2009. A global value chain approach to food safety and quality standards [R]. Durham: Duke University.
- [17] GRANE F G, 1984. Insurance principles and practices[M]. New York: Wiley.
- [18] MOWBRAY A H, Blanchard R H, 1995. Insurance: Its theory and practice in the United States [M]. New York: McGraw-Hill.
- [19] PORTHIN M, ROSQVIST T, Kunttu S, 2018. Risk assessment using group elicitation: Case study on start-up of a new logistics system[J]. Elicitation. Springer, Cham, 261: 511-527.
- [20] ROSENB J S, 1972. A case study in risk management[M]. New York: Meredinth Corp.
- [21] WANG J, YUE H, 2017. Food safety pre-warning system based on data mining for a sustainable food supply chain[J]. Food Control, 73: 223-229.
- [22] WILLAMS C A, Heine R M, 1985. Risk management and insurance[M]. New York: McGraw-Hill.

Elicitation of Big Data Technology on Market Supervision: Data-Driven Quality Risk Assessment

Zheng Liwei^{1,2,3}, Liu Zhonghua^{2,3}, Ni Chao^{2,3}

(1. Department of Industrial Engineering, Tsinghua University, Beijing 10084, China;

2. China Aero-Polytechnology Establishment, Beijing100028, China;

3. Key Laboratory of Quality Infrastructure Efficacy Research, SAMR, Beijing100028, China)

Abstract: The difficulty in obtaining quality information has always been a major challenge for government quality governance. The development of big data technology also makes it possible to collect, process and dynamically update multi heterogeneous data. This article focuses on quality risk assessment based on big data, and proposes a data-driven quality risk assessment idea, systematically identifies the risk sources of five different modes: supplying risk, running risk, admittance risk, credit risk and technology risk. Through the scattered data in the Internet, it constructs the indicators of different modes of risk sources, and finally designs the data-driven product quality risk assessment system framework and verified by an example of product defect risk early warning. The research here provides useful Elicitation for the application of big data technology in quality risk assessment.

Keywords: quality risk; data-driven; risk assessment; quality governance