

国际贸易进口食品准入风险分析技术研究

苏杨¹ 张彦彬¹ 刘丹² 林俊伟¹ 杜利民³ 郭学文⁴ 葛晓瑜³ 谷玥婵^{5*} 巩兴盛³

1 广州海关技术中心 2 中国电子口岸数据中心广州分中心 3 北京信城通数码科技有限公司

4 南京大学 5 华检通(北京)科技有限公司

DOI:10.12238/jmsr.v1i1.6156

[摘要] 随着我国进口食品的数量和品种逐年增加,不断发生的进口食品安全问题凸显了我国食品安全监管面临的挑战。针对国际贸易进口食品(乳制品、水产品、肉制品、酒精饮料)申报受理、现场查验、抽样送检和后续监管环节全链条,采用食品风险评估模型、知识图谱等信息技术,构建科学高效、智慧适用的进口食品风险评估模型和食品安全风险动态知识图谱,开展进口食品准入风险分析技术研究,助力提升我国国际贸易进口食品准入风险前置、预警与监管处置能力,缩短口岸通关时间,保障我国消费者健康。

[关键词] 国际贸易; 风险分析; 知识图谱; 进口食品风险评估模型

中图分类号: F74 **文献标识码:** A

Research on Risk Analysis Techniques for Import Food Access in International Trade

Yang Su¹ Yanbin Zhang¹ Dan Liu² Junwei Lin¹ Limin Du³ Xuewen Guo⁴ Xiaoyu Ge³

Yuechan Gu^{5*} Xinsheng Gong³

1 Guangzhou Customs Technology Center 2 China Electronic Port Data Center Guangzhou Branch

3 Beijing Itownet Cyber Technology Ltd 4 Nanjing University 5 Huajiantong (Beijing) Technology Co., Ltd

[Abstract] With the increasing number and variety of imported food in China, the continuous occurrence of imported food safety issues highlights the challenges faced by China's food safety supervision. For the whole chain of application acceptance, on-site inspection, sampling and submission for inspection and subsequent supervision of imported food (dairy products, aquatic products, meat products, alcoholic beverages) in international trade, the information technology such as food risk assessment model and Knowledge graph is used to build a scientific, efficient, intelligent and applicable risk assessment model for imported food and a dynamic Knowledge graph of food safety risk, and carry out technical research on risk analysis of imported food access, Assist in enhancing China's ability to anticipate, warn, and supervise the entry risks of imported food in international trade, shorten port clearance time, and ensure the health of Chinese consumers.

[Key Words] International trade; Risk analysis; Knowledge graph; Risk assessment model for imported food

1 研究背景

随着我国进口食品的数量和品种逐年增加,不断发生的进口食品安全问题凸显了我国食品安全监管面临的挑战。据海关总署公布的进口食品安全风险预警数据显示,2023年1-5月共从60个国家(地区)检出不符合我国法律法规和标准、未准入境食品809项、6630吨,主要不合格产品:饮料类、糕点饼干类、糖类、粮谷及制品类、酒类,不合格主要原因包括:品质不合格、证书不合格、标签不合格、食品添加剂超范围或超限量使用、微生物污染等。

经过多年的努力和探索实践,我国已经建立起一套以“预防在先、风险管理、全程监控、国际共治”的进口食品质量安全监管制度,覆盖了进口食品安全监管的所有相关环节,有力地保

障了进口食品安全。我国虽已积极建设食品安全风险分析机制,然而由于起步较晚,与一些发达国家相比,食品安全监管与风险分析技术还不够成熟。美国、日本、欧盟国家对进口食品安全监管开始较早,相应的监管体制较为完善,也最具代表性。美国、欧盟国家制定一套严格的监管体系,涵盖检验检疫系统、预警机制、风险评估机制等方面,大大降低食品安全风险。美国以风险管理为着眼点,推出“危害分析和关键控制点(HACCP)”食品安全监管模式^[1];欧盟通过建立HACCP体系实现对食品的溯源性管理,建立食品可追溯制度和统一联网的数据库^[2]。日本在食品安全监管方面所采取的保障体系和监管手段是世界上最为严格的,2006年即开始实行“肯定列表制度”,对化学污染物的限量标准进行了严格的限制^[3]。我国食品安全监管体系与上述发达国家相比存在一

定差距, 突出问题表现在: 检验监管大数据共享困难, 无法高效利用; 检测标准繁多杂乱, 检测方法与实验室资源不匹配、适用性差; 安全侦查与风险监控实验室建设及执行标准不统一, 这些问题在一定程度上制约了我国食品国际贸易的发展。

目前我国在进口食品安全准入管理方面的风险预警研究和实践相对较少。主要的风险信息数据来源主要包括国外TBT/SPS信息, 国外海关退货信息, 国外区域或国家风险预警信息, 社会反馈信息, 缺陷产品召回信息、产品伤害监测信息, 各行业组织、技术机构、行业性网络论坛等提供的第三方质量安全检测报告、质量安全风险信息, 信息掌握较为局限, 且具有一定的滞后性, 缺少贯通于条块、融通于各监管部门及社会的风险信息归集渠道。此外, 系统性安全风险问题发现机制仍需完善, 目前来说我国进口食品安全风险问题的发现主要借助于抽查检查、社会举报、国外预警信息跟踪等, 借助于大数据手段、通过对安全风险数据的智能分析, 提前发现行业性质量问题的技术手段尚未得到深入应用。

本文针对国际贸易进口食品(乳品、水产品、肉类)申报受理、现场查验、抽样送检和后续监管环节全链条, 采用食品风险评估模型、知识图谱等信息技术。通过分析进口食品安全相关数据以及社会舆情等相关信息, 构建科学高效、智慧适用的进口食品风险评估模型和食品安全风险动态知识图谱。

通过该模型识别出高风险的食品品类、生产企业、出口国家/地区等相关信息, 为相关部门进行科学有效的进口食品准入安全管理提供数据分析支撑, 整体提高我国的进口食品安全管理水平。基于知识图谱的食品安全查询与预警系统能实时为用户提供食品风险关联性查询服务, 为食品安全监控提供有力支撑。

2 国际贸易进口食品准入风险分析研究内容

本文研究内容包括3点, 分别是异构数据的获取、构建食品风险评估模型和知识图谱的构建与查询系统的研发准备, 如图1所示。



图1 研究内容

具体方法步骤如下:

(1) 基于大数据技术与机器学习算法, 通过分析公开的进口食品申报数据、合格评定数据、实验室检验数据、相关机构/部门的风险警示通报信息以及社会舆情等相关安全风险信息, 采用Hadoop/Spark架构以及ETL工具, 实现对多源、多类型数据的实时或准实时的采集, 并采用HBase分布式数据库对海量多类

型数据的存储检索; (2) 应用python作为大数据建模的基础工具, 以scikit-learn/tensorflow等机器学习/深度学习框架为基础, 搭建科学有效的进口食品风险评估模型, 通过该模型识别出高风险的进口食品品类、食品生产企业、出口国等相关信息, 协助相关部门进行科学有效的进口食品准入安全管理, 整体提高我国的进口食品安全管理水平; (3) 构建食品安全知识图谱, 在具备条件的情况下研发查询与预警模拟系统, 整个过程包括: ①知识抽取, 从一些公开的半结构化、非结构化的数据中提取出实体、关系、属性等知识要素; ②知识融合, 可消除实体、关系、属性等指称项与事实对象之间的歧义, 形成高质量的知识库; ③知识推理, 在已有的知识库基础上进一步挖掘隐含的知识, 从而丰富、扩展知识库。可基于已构建的知识图谱研发查询与预警系统, 提供实时查询服务。

2 国际贸易进口食品准入风险分析技术要点

2.1 应用态势感知模型构建食品风险评估模型

整体流程如图2所示, 首先根据报关信息, 搜索准入名册数据库(数据库一)判断是否在准入名册, 不在准入名册的拒绝申报并预警, 在准入名册时, 运用基于舆情信息的综合评价模型, 计算该批次食品总体风险值, 确定总体风险等级, 判断是否抽中布控, 未被抽中可通关放行, 被抽中则运用基于实验室检测数据的综合评价模型(有实验室数据时)或基于AHP-模糊综合评价法的风险评价模型(无历史数据时), 进行抽检项目风险等级计算, 确定检测查验/送检项目。

根据检测结果综合评定是否合格, 合格的可通关放行, 不合格的采取相应处置和预警。

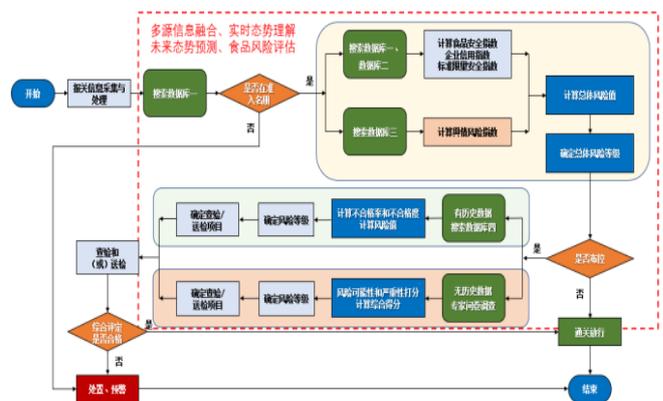


图2 态势感知模型实现流程

数据库一

(1) 准入国家、企业名单; (2) 全球食品安全指数排名; (3) 进口企业诚信记录。

数据库二

(1) 国内法律法规、标准; (2) 国外法律法规、标准。

数据库三

(1) 国际舆情通报; (2) 国内市场监督管理抽检。

数据库四

(1) 口岸抽检检测; (2) 风险监测检测。

根据食品特性、企业特性、国家特性、时间特性等进行风险识别,建立了风险评价指标体系,其中食品特性下的指标是有实验室检测数据的,利用这部分可以建立基于实验室检测数据的评价模型,对于无历史数据的指标可以建立基于AHP-模糊综合评价法的风险评价模型,而企业特性、国家特性等无法通过实验室检测数据进行评价,这部分基于舆情信息挖掘建立基于舆情信息的评价模型。

三个评价模型构成最终的食品风险评价模型。

2.2知识图谱的构建

本文采用资源描述框架语言(Resource Description Framework,简称RDF)结构来设计仪器设备知识结构模型。RDF表述的资源都通过主语、谓语、宾语这三元组来描述的,分别表示资源、资源的属性、资源的值。

2.2.1知识抽取

(1)数据采集和处理。仪器设备输出的文档,主要有以下几种格式: csv,txt,excel,doc,pdf。其中csv/excel/txt格式可视为半结构化文档,经过简单转换即可转化为结构化数据。Doc(word格式),经过简单的export也可以转化为txt文件,进而从中提取结构化数据。大多数仪器设备,由于需要同时输出文字、表格、图表,广泛采用了pdf格式作为输出文档格式,同时pdf格式作为一种面向呈现的格式,其结构和读取方式比较复杂,因此本文主要关注pdf格式文档的读取。本文应用PDFMiner^[4]工具从PDF文档中提取信息,获取页面中文本的确切位置,以及其他信息,例如字体或线条。(2)选择研究设备。本文选择一种实验室较为常用的HPLC2设备输出的PDF文档来进行数据分析研究,通过收集一些样本在HPLC2下的实验结果,我们获得实验人员从仪器设备的输出文档中最为关注的部分。通过读取pdf文档中的文本内容及对应的数据结果,并将它们关联到仪器输出文档,我们建立了有监督的机器学习的数据集:原始特征是设备输出pdf文档中的文本,标签是数据结果。(3)模型构建。本文采用命名实体识别(Named Entity Recognition,简称NER),又称作“专名识别”,识别文本中具有特定意义的实体,例如人名、地名、机构名、专有名词等,从文本信息中提取关键信息。然后采用了BIO标注法对文本信息进行标注,并应用Bert-BiLSTM-CRF^[5]架构对模型进行训练,实现对文本的智能分析。

2.2.2知识图谱存储

本文设计三元组表,每条三元组存储为一行具有三列的记录,如图3所示,将知识图谱的RDF映射到关系数据库。

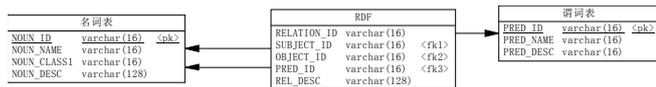


图3 三元组表

2.2.3知识图谱的可视化

基于JAVAEE平台上构建知识图谱可视化平台,平台使用了多层的开发架构,其中DAO负责数据访问,Service、Manager层负

责业务的逻辑访问,Controller负责视图展示层的数据准备。View层可以为JSP,HTML或手机端,大部分采用模板与JSON数据进行界面渲染。

基于以上的平台架构,可开发出知识图谱的可视化功能,具体效果如图4所示:

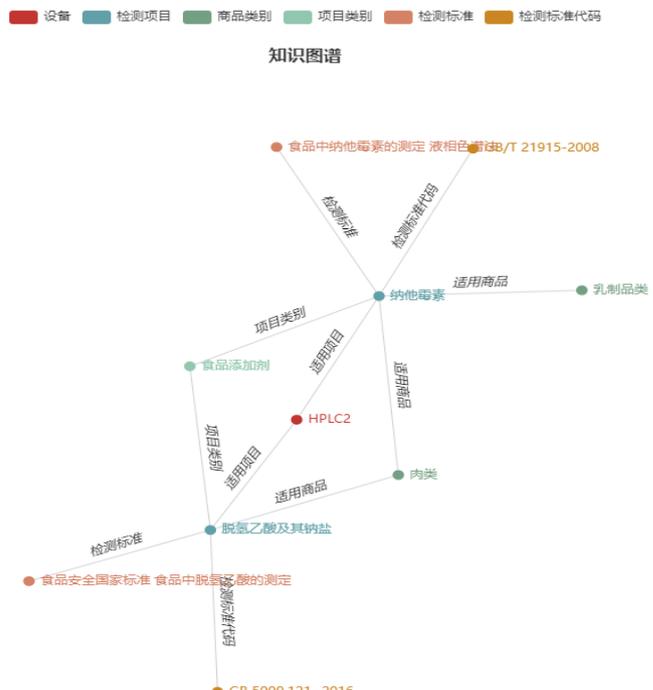


图4 知识图谱可视化展示

3 结束语

本文探索了将食品风险评估模型和知识图谱这两项信息技术应用于国际贸易进口食品准入风险分析,建立了科学有效的进口食品风险评估模型,通过该模型识别出高风险的进口食品品类、食品生产企业、出口国等相关信息,构建食品安全知识图谱,实现进口食品风险数据的可视化展示,协助相关管理部门进行科学有效的进口食品准入安全管理,整体提高我国的进口食品安全管理水平。

[参考文献]

[1]杜琨,张亚宁.危害分析和关键控制点(HACCP)原理在食品工业中的应用[J].安徽农业科学,2005,(05):881-882.
[2]庞乐君,李哲.全球食品溯源法规制度现状和比较[J].上海预防医学,2015,27(06):305-307.
[3]臧敏,季任天.日本肯定列表制度探析——以食品安全标准为视角[J].标准科学,2010,(03):77-82.
[4]PdfMinerdocs,https://pdfminer-docs.readthedocs.io/pdfminer_index.html.
[5]谢腾,杨俊安,刘辉.基于 BERT-BiLSTM-CRF 模型的中文实体识别[J].计算机系统应用,2020,29(07):48-55.