

# 基于深度置信网络-多类模糊支持向量机的粮食供应链危害物风险预警

王小艺<sup>1,2</sup>, 李柳生<sup>1</sup>, 孔建磊<sup>1,2,\*</sup>, 金学波<sup>1,2</sup>, 苏婷立<sup>1</sup>, 白玉廷<sup>1</sup>

(1.北京工商大学计算机与信息工程学院, 北京 100048;

2.北京工商大学北京市食品安全大数据技术重点实验室, 北京 100048)

**摘要:** 近些年粮食供应链安全问题频发, 为减少食源性风险威胁, 风险预警正逐渐成为强化粮食食品安全体系的有力保障。但现有风险预警方法在面对多源异构非结构化食品数据时, 存在预警准确率低、人工成本高等局限问题。本文在分析全国26个省份的大量抽检数据及关联信息基础上, 建立了基于深度置信网络(deep belief network, DBN)-多类模糊支持向量机(multiclass fuzzy support vector machine, MFSVM)的风险分级预警模型, 先对海量粮食供应链抽检数据进行嵌入编码和归一化处理, 获得结构化食品数据; 将其输入到DBN模型进行高维度特征提取, 自适应地挖掘供应链中各危害因素间风险变化及内在关联概率, 最后将高维特征输入到优化的MFSVM进行训练, 实现供应链中各主要危害物风险分级预警。对比实验结果表明, DBN-MFSVM模型在粮食抽检数据上具有更好鲁棒性和泛化性, 其准确率达到98.44%, 运行时间85 s, 可快速识别出粮食供应链中危害物风险程度和优先次序, 为监管部门制定有针对性的抽检策略、确立优先监管领域和分配风险监管资源提供科学依据。

**关键词:** 粮食供应链安全; 风险预警; 深度置信网络; 多类模糊支持向量机

## Risk Pre-warning of Hazardous Materials in Cereal Supply Chain Based on Deep Belief Network-Multiclass Fuzzy Support Vector Machine (DBN-MFSVM)

WANG Xiaoyi<sup>1,2</sup>, LI Liusheng<sup>1</sup>, KONG Jianlei<sup>1,2,\*</sup>, JIN Xuebo<sup>1,2</sup>, SU Tingli<sup>1</sup>, BAI Yuting<sup>1</sup>

(1. School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China;

2. Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China)

**Abstract:** In recent years, food supply chain security problems have occurred frequently. In order to reduce the threat of foodborne risks, risk early warning is becoming a powerful guarantee to strengthen the food safety system. However, in the face of multi-source heterogeneous unstructured data on foods, the existing risk warning method is limited by its low early warning accuracy and high labor cost. Based on analysis of a large number of data from food sample survey and related information from 26 provinces in China, this paper establishes a risk classification and early warning model using deep belief network-multiclass fuzzy support vector machine (DBN-MFSVM). First, the numerous data from grain supply chain sample survey are embedded, coded and normalized to obtain structured food data, which are then input into the DBN model to extract high-dimensional features, and self-adaptively mine the risk change and intrinsic correlation probability among the risk factors in the supply chain. Finally, the high-dimensional features are input into the optimized MFSVM for training to realize the risk classification and early warning of the main hazards in the supply chain. The comparative experimental results show that the DBN-MFSVM model has better robustness and generalizability for grain sample survey data. Its accuracy rate is 98.44%, and the running time is 85 s. It can quickly identify the risk level and priority of hazardous materials in the food supply chain, and thereby provide a scientific basis for the regulatory authorities to develop targeted sampling strategies, establish priority regulatory areas and allocate risk monitoring resources.

**Keywords:** food supply chain security; risk early warning; deep belief network; multiclass fuzzy support vector machine

DOI:10.7506/spkx1002-6630-20190916-205

中图分类号: TS210.1; TP183

文献标志码: A

文章编号: 1002-6630(2020)19-0017-08

收稿日期: 2019-09-16

基金项目: “十三五”国家重点研发计划重点专项(2017YFC1600605); 北京市教育委员会科技计划一般项目(KM201910011010)

第一作者简介: 王小艺(1975—)(ORCID: 0000-0003-1047-4245), 男, 教授, 博士, 研究方向为粮油供应链风险预警与防控、智能感知与优化。E-mail: sdwangxy@163.com

\*通信作者简介: 孔建磊(1989—)(ORCID: 0000-0002-0074-3467), 男, 讲师, 博士, 研究方向为数据挖掘及融合、粮油供应链风险评估及预警。E-mail: kongjianlei@btbu.edu.cn

引文格式:

王小艺, 李柳生, 孔建磊, 等. 基于深度置信网络-多类模糊支持向量机的粮食供应链危害物风险预警[J]. 食品科学, 2020, 41(19): 17-24. DOI:10.7506/spkx1002-6630-20190916-205. <http://www.spkx.net.cn>

WANG Xiaoyi, LI Liusheng, KONG Jianlei, et al. Risk pre-warning of hazardous materials in cereal supply chain based on deep belief network-multiclass fuzzy support vector machine (DBN-MFSVM)[J]. Food Science, 2020, 41(19): 17-24. (in Chinese with English abstract) DOI:10.7506/spkx1002-6630-20190916-205. <http://www.spkx.net.cn>

国以民为本, 民以食为天。作为人类赖以生存的重要商品, 粮食安全问题是关系国计民生的战略问题, 既影响着国家主权、经济和社会的稳定发展, 又影响着人民身体的健康。然而近些年粮食安全问题时有发生, 粮食供应从原材料到最终商品, 需要经过种植收购、生产加工、仓储运输、销售消费等多个环节, 是一个是多维、复杂、耦合的过程<sup>[1]</sup>。其中任何一个环节都存在不同类别和程度的危害物风险因素, 且每个因素受到食品多样性、数据多源异构、地区分布差异、时间变化性等因素影响, 这直接导致粮食供应链安全得不到保障, 自然产生和非法添加的食源性危害物如重金属、农药残留、食品添加剂以及卫生条件控制不当造成的微生物、真菌毒素污染, 严重威胁着群众的身体健康和生命安全<sup>[2-3]</sup>。为有效控制粮食供应链中每个环节的潜在污染源, 预防和减少食源性风险威胁, 风险预警正逐渐成为强化粮食供应链安全体系的有力保障<sup>[4]</sup>。所谓预警是对潜在危害物及可能产生风险信息进行动态监控、风险评估和综合分析判断的结构化决策过程, 在全面收集预警指标数据、汇总多因素风险信息基础上, 通过对相关数据分析、信息归纳、知识推演等处理, 去伪存真, 判断事物发展的走势, 进行合理预测, 提出不同警情的对策建议和反馈<sup>[5]</sup>。构建合理有效的粮食供应链风险监测与预警体系, 在最大范围内有效防控风险, 降低食品安全事故的发生概率, 已成为当前亟待解决的问题。

目前国内外学者在食品风险预警领域做了很多研究, 总体而言, 食品安全预警系统可分为信息采集系统、预警分析系统、应急防控系统3个主要部分, 其中预警分析系统是整个食品安全风险预警系统的关键和核心。典型如欧盟批准的食品饲料快速预警系统, 在危害分析关键控制点标准下构建, 重点注重最相关的高风险危害物。而世界卫生组织建立的全球环境食品污染监测和评估系统, 对全球26个国家进行长期不间断微生物危害事件及疫情监控、动态监测和预警潜在危害风险<sup>[6]</sup>。近些年, 我国食品安全监测和防控系统也在迅速发展。农业农村部通过无公害农业与食物安全预警系统对食源性农药残留、重金属超标监测, 向全国发布农产品警情。国家市场监督管理总局为有效应对食品安全问题建立了

包括信息处理系统、预警分析系统和快速反应系统在内的预警体系<sup>[7]</sup>。国家食品安全风险评估中心构建了涵盖全国31个省份的全国食品风险监测和预警系统, 涵盖生物性危害(细菌、病毒和寄生虫)、化学危害和微生物等各类危害物, 开展食品暴露监测及食源性疾病监测和调查工作<sup>[8]</sup>。虽然目前中国已建立诸多食品安全风险监测与预警防范体系, 但仍存在较多问题, 如风险监测与预警脱节、供应链间危害因素迁移规律不清等, 特别是预警方法不先进问题更为突出, 导致难以做到与食品监测信息系统的连接, 无法真正地做到对出现的食品安全风险提前预报<sup>[9]</sup>。

目前食品风险预警方法主要分为定性分析<sup>[10]</sup>与定量数据挖掘<sup>[11]</sup>两种。定性分析方法依托从调查问卷、专家咨询、政策与理论分析中得到数理统计数据, 抽取关键危害因素及潜在关联, 在食品管制、食品安全风险态度分析、食品安全问题认知度等研究中很常见; 但其依赖的数据存在主观性, 掩盖了危害物在供应链环节内耦合作用, 容易导出违背事实规律的伪结论。而基于定量数据挖掘的预警方法, 从大量的食品安全监督抽检数据入手, 在生产、流通以及消费环节采集超标数值、不合格率等信息, 实现食品安全隐患评价和预测预警<sup>[12]</sup>。常见数据挖掘方法有贝叶斯网络<sup>[13]</sup>、关联规则<sup>[14]</sup>、支持向量机<sup>[15]</sup>、决策树<sup>[16]</sup>、误差逆传播(back propagation, BP)神经网络<sup>[17]</sup>等。Wang Jing等使用模糊神经网络模型来预测并描述食品在各环节中的微生物数量, 并以此作为参考来预警微生物风险, 实现食品安全动态可追溯监管<sup>[18]</sup>。Wang Xueli等根据国家食品安全风险评估公布的湖南省21 866例食源性疾病, 就诊疗病例中确诊病例21 226例这一数据, 构建了贝叶斯网络模型预测每日真实患者数<sup>[19]</sup>。Zhu Changxing等为了克服传统食品风险预警模型在假设、样本量、泛化能力等方面的不足, 将支持向量机分类理论与预警理论相结合, 提出了一种食品安全预警模型<sup>[20]</sup>。王霞利用2010年食物摄入量数据构建了BP神经网络风险预警模型, 预测2011年的食物摄入量<sup>[21]</sup>。这些方法科学有效地把监督抽检的数据转化为风险的量化分级, 但粮食供应链中环节众多、危害物来源各异, 现有数据挖掘方法需要依赖大量人工干预实现

指标量化和权重分配,不利于日常食品安全预警应用和管理<sup>[22]</sup>。而且已有研究只是在当前数据出现超标、数据变动量超标时根据统计值做出预警,适用于小数据量预警处理,只在风险发展到临界点或即将到临界点时才呈现警示结果,不具备早期预测未来风险的预警本质,无法从海量多维异构数据中挖掘粮食供应链危害物风险预警作用机理,导致预警准确率低、人力工作量大、监管实施难等问题<sup>[23-24]</sup>。

相比传统预警方法存在的受供应链环节、危害物迁移扩散等众多因素的影响而经常呈现出非线性、波动性大等问题,深度置信网络(deep belief network, DBN)应用多层次结构模拟大脑认知神经系统,将数据间复杂关系抽象为特征或属性类别等高层表示<sup>[25]</sup>,迅速得到了广大学者的高度关注,其通过大样本训练获得系统隐含规律,不需要严格的输入值间、输入输出值间假设关系,同时能够以区间数、模糊数等方式处理定性信息,目前DBN在产量预测<sup>[26]</sup>、水稻病害预警<sup>[27]</sup>、储粮虫害预警<sup>[28]</sup>、药品鉴别<sup>[29]</sup>等方面应用也得到了充分的肯定,能够在系统内部状态未知条件下实现基于现有定量定性知识的状态预测、预警。因此,针对粮食供应链危害物风险分级预警问题特点,本研究在分析中国质监部门实际抽检数据及预处理方法基础上,结合多类模糊支持向量机(multiclass fuzzy support vector machine, MFSVM)和DBN模型优势,自适应分析粮食供应链中多维异构抽检数据,构建基于DBN-MFSVM的粮食供应链危害物风险分级预警模型,从而达到充分掌握供应链内各类危害因素迁移变化规律的目的,使危害物风险预警更加准确。

## 1 材料与方法

### 1.1 数据来源与特点

本研究以全国主要粮食生产地区的主要危害物进行实例分析,搜集整理国家粮食局、国家市场监督管理总局等网站公布的抽检数据,按照重金属、霉菌毒素、微生物、食品添加剂、农药残留等危害物类别及供应链环节来源进行分类整理和预处理,收集整理到2013—2018年的粮食及其加工品的抽检数据,包括除港澳台、西藏、新疆、内蒙古、宁夏和甘肃在外的26个省份,涵盖中国粮食食品的主要产区 and 消费大省,亦属于人口密集地区。粮食食品种类由大米、大米加工品、小麦粉、小麦粉加工品、其他粮食加工品组成,这5类产品涵盖了我国消费者最主要的粮食结构。抽检数据整体反映了各省份地区粮食供应链中危害物的类别、来源以及有关危害物的信息,每个数据样本由多个因素组成,包括:产品

名称、标称生产企业及被抽样企业信息、抽样环节及场所、生产及抽检日期、食品分类、抽检项目及结果、标准值等。

鉴于粮食供应链安全受到社会、经济、监管、危害物毒性等众多因素影响,且供应链涵盖从种植、生产加工、仓储物流等众多环节,以及含各类商场、超市及其他消费场所在内的销售消费环节,这些环节均可能出现安全隐患和风险因素。因此,本研究深入整理收集粮食供应链相关信息,通过对文献、食品行业专业网站、新闻媒体信息的分析,得到匹配食品抽检数据的粮食供应链主要危害物计量统计信息。其中,参考国家统计局的第六次全国人口普查结果和《中国统计年鉴》获取粮食相关人口数据和消费数据;致病菌的污染情况、危害程度和流行病学数据则来源于《中国卫生统计年鉴》、《中国食品工业年鉴》。此外,其他信息还包括调研信息,包括统计中国食品安全网等网站公告的粮食安全事件情况,并采用调查问卷的方式,共邀请来自高等院校、食品药品监管部门、食品行业的专家及从业人员进行调研论证,发出共回收有效问卷553份,获取蕴含专家综合分析及预警知识的调研问卷信息。

本研究在分析不同省份、不同时间的食品抽检数据基础上,结合计量统计信息和调研问卷信息辅助评价,构建多维层次指标体系,对粮食供应链危害物进行分级量化分析,初步评定出各类危害物风险等级。本指标体系分为14个一级指标和34个属性指标,如表1所示,每个一级指标由其多个属性指标加权求和得到,而上层评价指标需通过下层评价指标的评价结果反映出来,即涵盖定性指标,如社会关注度、危害程度、监管可及性等,同时兼顾定量指标,如全省年度总产量、粮食生产及消费价格、国家标准限定值等内容。各指标权重赋值由高等院校、食品药品监管部门、食品行业的专家及从业人员综合评价得出,伴随调研及问卷实施过程获得。根据专家们综合评定,对各样本数据中危害物进行风险分级,共细分出8个等级:安全、较安全、预警、较低危险、中危险、较高危险、高危险和超高危险<sup>[30]</sup>。综合抽检数据、计量统计数据、调研数据等构建多维异构数据集,本研究在分析多维异构食品抽检数据的基础上,将粮食供应链划分为生产、流通、消费3个主要环节,以粮食供应链中的重金属(镉、铅、汞等)、霉菌毒素(黄曲霉毒素、脱氧雪腐镰刀菌烯醇、赭曲霉毒素、玉米赤霉烯酮等)、农药残留(马拉硫磷、戊唑醇等)以及其他主要危害物为研究对象,这些危害物反映了粮食供应链中主要危害物大类,会从各个方面影响人的身体健康。



表1 粮食供应链危害物数据结构

Table 1 Hazard data structure sampled in cereal supply chain

一级指标	属性指标及特征描述	特征变量
食品类别	大米、小麦粉、面食、米线、零食制品等	$x_1$
抽检省份	北京、黑龙江、天津等26个省市地区	$x_2$
样品类别	大米、大米加工品、小麦粉、小麦粉加工品、其他粮食加工品等	$x_3$
抽样环节	生产、流通、消费	$x_4$
区域类型	区域(城市、乡村)、抽检场所(超市、农贸市场等)	$x_5, x_6$
生产日期	年份(2013—2018)、月份(1—12)、日期(1—31)	$x_7, x_8, x_9$
抽检日期	年份(2013—2018)、月份(1—12)、日期(1—31)	$x_{10}, x_{11}, x_{12}$
抽检项目	危害物大类, 具体危害物类别, 如: 重金属、霉菌毒素、农药残留等	$x_{13}, x_{14}$
检验指标	检验值、国家标准限定值、超标倍数	$x_{15}, x_{16}, x_{17}$
单位	包括CFU/g、mg/kg、MPN/g、μg/kg等	$x_{18}$
危害性	致癌性、半数致死量、毒性、致病严重程度	$x_{19}, x_{20}, x_{21}, x_{22}$
社会性	焦虑度、总体关注度、百度PC指数、百度移动指数	$x_{23}, x_{24}, x_{25}, x_{26}$
经济性	全省年度总产量、全省年度总播种面积、生产价格、消费价格	$x_{27}, x_{28}, x_{29}, x_{30}$
监管性	监管可及性、危害程度、法律标准完备性、预警体系完整性	$x_{31}, x_{32}, x_{33}, x_{34}$
等级	安全等级I、较安全等级II、预警等级III、较低危险等级IV、中危险等级V、较高危险等级VI、高危险等级VII、超高危险等级VIII	$y$

经过上述过程,整合不同来源、不同类型的数据,形成粮食供应链危害物数据集,共有26 285个数据样本。将数据集划分为训练集合和测试集合,划分比例为80%:20%。所获得数据呈现多源头、多属性、强关联等特点,区别于图像、视频、文本等非结构化数据,每个数据样本由一行多列形式记录,存储在二维表中,除了整数、浮点等数值型定量特征,还包含非数字型定性特征,比如中文、英文字母、特殊符号等,是典型的多源异构的数据,涵盖了粮食供应链各环节中不同危害物潜在风险因素。

## 1.2 基于DBN-MFSVM的风险等级预警模型

本实验基于DBN-MFSVM的等级预警模型(图1)架构,其包含3个步骤:1)将粮食数据中的类别特征进行嵌入编码,再对数值型特征进行归一化处理,最终预处理为结构化数据。2)使用DBN对结构化数据进行特征提取,挖掘供应链危害物风险变化内在关联,形成出高维度表征的特征集合。3)以高维特征为输入,应用MFSVM进行分类器参数优化,准确、高效地实现各主要危害物风险分级预警。

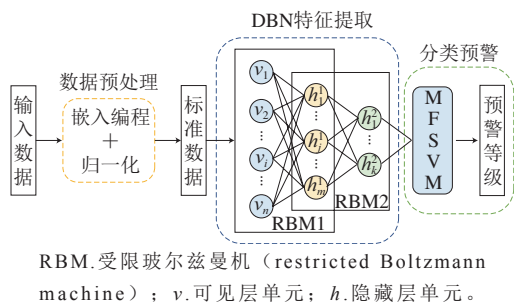


图1 DBN-MFSVM模型

Fig. 1 DBN-MFSVM model

### 1.2.1 基于嵌入归一化的数据预处理

输入 $n$ 个训练数据 $X_n = \{x_1, x_2, \dots, x_m\}$ 预警等级 $y_n$ 。在多维异构食品非结构化数据中,特征有数值型特征和中英文类别型特征。但计算机只能处理数值型数据,没法处理类别型数据。因此,需要对各种类别特征进行相应的编码,进行定性指标的定量化处理,本实验使用标签编码和独热编码相结合的编码方式,先应用标签编码将分类特征转化成连续的数值型变量,继而应用独热编码有效增加额外列属性,让0和1出现在对应的列分别表示每个分类值有或无。将样本中某个特征 $x_m$ 进行标签独热编码,如式(1)所示。

$$x_m \mapsto \delta_{x_m \alpha} \quad (1)$$

式中: $\delta_{x_m \alpha}$ 是克罗内克函数, $\alpha$ 与 $x_m$ 是两个输入,当 $\alpha = x_m$ 时 $\delta_{x_m \alpha}$ 为1,不相等时 $\delta_{x_m \alpha}$ 为0。如果 $N$ 是特征 $x_m$ 可能的数量,则 $\delta_{x_m \alpha}$ 是长度为 $N$ 的向量。举例说明,如抽检地分别为北京、山西、陕西,经过编码后表示为“北京100,山西010,陕西001”,这样消除了数值大小关系比较。但本数据中抽检地有26个省份地区,直接编码会导致维度急剧增加,产生大量冗余的稀疏矩阵,加重了预警系统计算量。因此,在本实验中采用嵌入方式进一步挖掘不同维度之间的内在关系,对特征进行神经网络嵌入,如式(2)所示。

$$X_m = \sum_{\alpha} \omega_{\alpha \beta} \delta_{x_m \alpha} = \omega_{x_m \beta} \quad (2)$$

式中: $\omega_{\alpha \beta}$ 是连接独热编码层与嵌入层的权重; $\beta$ 是嵌入层的索引,映射的嵌入只是这一层的权重。如此,所有特征被嵌入到一个低维空间,每一个 $n \times m$ 的矩阵 $M$ 都定义了 $\mathbf{R}^m$ 到 $\mathbf{R}^n$ 的一个线性映射: $x \mapsto M_x$ 。 $M_x$ 对应矩阵 $M$ 中的一行,如式(3)所示。

$$[0 \ 0 \ 1 \ 0 \ 0] \times \begin{bmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{23} \\ \omega_{31} & \omega_{32} & \omega_{33} \\ \omega_{41} & \omega_{42} & \omega_{43} \\ \omega_{51} & \omega_{52} & \omega_{53} \end{bmatrix} = [\omega_{31} \ \omega_{32} \ \omega_{33}] \quad (3)$$

使用嵌入来表示所有类别特征之后,所有嵌入层和所有连续变量的输入被连接起来,合并的图层被视为神经网络中的普通输入图层,可以使用标准反向传播方法训练整个网络,如此可以了解每个类别的内在属性,更深层地挖掘各特征组合的内在关联。鉴于各个特征属性定义及取值范围均不同,为了消除属性之间的量纲影响,需要对各属性进行归一化处理,使得不同属性之间具有可比性。对数值类型的特征进行线性函数归一化,将嵌入编码的数据对进行线性变换,将所有特征值范围定义在 $[0, 1]$ 内,实现对原始数据的等比缩放,为后续深度置信网络训练提供底层特征集合。归一化公式如式(4)所示。

$$X_{\text{norm}} = \frac{X - x_{\min}}{x_{\max} - x_{\min}} \quad (4)$$

式中:  $X$ 为原始样本特征值;  $X_{\text{norm}}$ 为归一化后的特征值,  $x_{\text{max}}$ 和 $x_{\text{min}}$ 分别为该特征中的最大值和最小值。

### 1.2.2 DBN特征提取

本实验所使用的DBN是由多层RBM和一层BP神经网络组成, 训练过程主要包含预训练和微调2个阶段。

预训练阶段: 使用对比散度算法, 逐层来调整连接权重和偏置, 首先训练输入层和隐藏层之间的参数, 把训练后得到的参数作为下一层的输入, 再调整该层与下一个隐藏层之间的参数, 然后逐次迭代, 完成多层RBM的训练。RBM由可见层和隐藏层构成的两层结构, 可见层和隐藏层又分别由可见变量和隐藏变量构成。本实验中的DBN由一个可见层 $v$ 和两个隐藏层 $h_1$ 、 $h_2$ 构成, 相邻的两层即是一个RBM, 则 $v$ - $h^1$ ,  $h^1$ - $h^2$ 构成了两个RBM, 具体见图1。

$v$ 表示所有的可见层单元, 其输入为嵌入归一化后的数据特征;  $h$ 表示所有隐藏层单元, 模型的3个参数包括权重矩阵 $W$ 、可见层单元偏置 $A$ 、隐藏层单元偏置 $B$ 。对于一个有 $n$ 个可见单元 $m$ 个隐单元的RBM模型,  $v_i$ 表示第 $i$ 个可见单元,  $h_j$ 表示第 $j$ 个隐单元, 它的参数形式为:  $W_{ij}$ 表示第 $i$ 个可见单元和第 $j$ 个隐单元之间的权值,  $a_i$ 表示第 $i$ 个可见单元的偏置,  $b_j$ 是第 $j$ 个隐单元的偏置。RBM是根据能量平衡原理的一种概率生成模型, 利用无监督学习初始化权重值, 从而学习出特征间的内在结构和关系。对于一组给定状态下的 $(v, h)$ 值, 观察数据得知可见层和隐藏层均服从伯努利分布, 得到RBM的能量函数(公式(5))。

$$E(v, h|\theta) = -\sum_{i=1}^n a_i v_i - \sum_{j=1}^m b_j h_j - \sum_{i=1}^n \sum_{j=1}^m v_i W_{ij} h_j \quad (5)$$

式中:  $\theta = (W_{ij}, a_i, b_j)$ , 为RBM模型的参数, 能量函数表示在每一个可见层节点和每一个隐藏层节点之间都存在一个能量值。对该能量函数指数化和正则化后可以得到可见层节点集合和隐藏层节点集合分别处于某一种状态下 $(v, h)$ 联合概率分布公式(式(6)), 其中 $Z(\theta)$ 按公式(7)计算。

$$P(v, h|\theta) = \frac{e^{-E(v, h|\theta)}}{Z(\theta)} \quad (6)$$

$$Z(\theta) = \sum_{v, h} e^{-E(v, h|\theta)} \quad (7)$$

式中:  $Z(\theta)$ 为归一化因子, 表示对可见层和隐藏层节点集合的所有可能状态的求和。由于RBM模型层间全连接、层内无连接, 在给定可见单元的状态时, 各隐藏层单元的激活状态之间是条件独立的。此时, 第 $j$ 个隐单元的激活概率可按式(8)计算。相应地, 当给定隐单元的状态时, 可见单元的激活概率同样是条件独立的。此时第 $i$ 个可见单元激活的概率可按式(9)计算。

$$P(h_j=1|v) = f(b_j + \sum_i v_i W_{ij}) \quad (8)$$

$$P(v_i=1|h) = f(a_i + \sum_j W_{ij} h_j) \quad (9)$$

式中:  $f$ 为激活函数, 这里选用sigmoid作为激活函数。是因为它可以把 $(-\infty, +\infty)$ 的值映射到 $[0, 1]$ 这

个区间。也就是说, 无论模型的可见层输入节点数据处于一个多大的范围内, 都可以通过sigmoid函数求得它相应的函数值, 即节点的激活概率值。

微调阶段: 把训练RBM得到的输入作为BP神经网络的输入, 进行前向传播, 然后进行反向传播, 将顶层训练的误差从输入层反向传播至输入层, 微调各层的参数, 使得误差最小。最终将DBN特征提取后的特征作为MFSVM分类器的输入, 实现预警等级的分类。

### 1.2.3 多类模糊支持向量机

传统的支持向量机的两类性较强, 但对本实验涉及的多类安全等级问题则表现欠缺。因此本实验使用模糊输入和模糊输出的支持向量机作为分类器, 以径向基函数描述核函数, 优化隶属度表现形式, 并使用有向无环图(directed acyclic graphs, DAG)的形式将双模糊支持向量机推广到多类问题, 构建MFSVM, 提升支持向量机的分类能力和泛化性。将DBN模型提取的高维特征作为MFSVM分类器的输入, 取 $n$ 个训练样本的高维特征及标签构建集合 $\{(h_i^2, y_i), i=1, 2, \dots, n\}$ , 通过模型训练在样本空间中划分超平面, 将不同类别的样本分开, 其目标函数见式(10)。

$$\min L(w, \zeta) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \zeta_i \quad (10)$$

$$\text{约束条件: } y_i (w^T x_i + b) \geq 1 - \zeta_i, i=1, \dots, n$$

式中: 惩罚参数 $C$ 是用来限制奇异点存在的程度;  $w$ 和 $b$ 分别为超平面的法向量和截距; 松弛变量 $\zeta_i$ 对一些不能被超平面正确分类的样本进行补偿。之后, 将拉格朗日算子 $\alpha_i$ 引入公式, 通过化简和转化, 可以得到超平面的表达式(式(11))。

$$f(x) = \text{sgn} \left( \sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right) \quad (11)$$

式中:  $K(x_i, x)$ 是为了解决非线性问题引入的核函数。选择了常用的径向基函数作为核函数,  $\sigma$ 为核函数的带宽, 控制函数的作用范围, 其表达式见公式(12)。

$$K(x_i, x) = \left( -\frac{\|x_i - x\|^2}{2\sigma^2} \right) \quad (12)$$

实际应用中, 每个训练样本对支持向量机所起的作用不同, 边缘数据最容易错分且成为支持向量的机会较多, 而中间数据成为支持向量的概率较小。因此在分类超平面求解过程中, 需要引入模糊理论对不同的训练样本分布情况进行描述, 定义了一个与分类超平面垂直的隶属度函数 $s_i$ , 表达每个样本对分类超平面的重要性, 将多个隶属度函数组合在一起可减少不可分区域, 如此构建模糊支持向量机, 完成目标函数的惩罚参数 $C$ 和松弛变量 $\zeta$ 的平衡, 目标函数(式(10))可改写为式(13)。

$$\min L(w, \zeta, s) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n s_i \zeta_i \quad (13)$$

约束条件:  $y_i (w^T x_i + b) \geq 1 - \zeta_i, i = 1, \dots, n$

式中:  $s_i$  的取值区间为  $(0, 1]$  表示了每个样本属于标签的可能性, 该变量的取值较小, 从而对惩罚参数  $C$  的取值进行了一定程度的限制, 防止其取值过大或过小对分类性能产生影响, 使分类器的性能更加稳定。在模糊向量机中时, 每个样本的权值是通过隶属度函数来求得的, 那么隶属度函数的选择和设计是模糊向量机模型建立的关键。本实验采用的隶属模型是基于样本到类中心之间的距离来度量其隶属度的大小。随后, 在上述基础上, 将模糊支持向量机向多类问题的推广, 使用 DAG 方法将多个两类分类器组合成多类分类器。在训练阶段, 对于本研究  $k$  类安全等级预警问题, 设计  $k(k-1)/2$  二类分类器。而在决策阶段, 使用从根节点开始的导向非循环图, 具有  $k(k-1)/2$  个内部节点以及  $k$  个叶子节点, 每个内部节点都是一个二类分类器, 叶子节点为最终的类值。对一个测试样本, 从根节点开始根据分类器的输出值决定其走左侧或右侧路径, 一直到叶子节点为止得到样本所属的类值, 最终构建了 MFSVM 模型进行安全等级的预警。

#### 1.2.4 模型参数选择

DBN 中的“深度”体现在有多个 RBM 层, 隐藏层越多, 网络能更加深度挖掘数据, 但也导致网络计算耗时越大, 效率越低。本实验选择由 2 个 RBM 堆叠而成的 DBN, 以及一个 MFSVM。DBN 模型的节点数分别为 83、70、50。其中, 83 个输入变量表示输入的抽检数据的维度, 包含了各个属性, RBM 的神经元个数分别为 70、50。训练中为了避免过拟合, 设置迭代次数为 100, 学习率为 0.001。

惩罚参数  $C$  和松弛变量  $\zeta$  的选取对 MFSVM 模型的性能影响较大, 目前主要采用经验确定法和网格搜索法, 本实验采用网格搜索法对 MFSVM 模型进行参数选择, 用优化后的 MFSVM 模型对危害物安全等级进行识别并输出预警结果。惩罚参数  $C$  代表错误的容忍程度, 核参数  $\sigma$  是核宽度。过高或过低的参数值会造成 MFSVM 模型的“过学习”或“学习不足”现象, 造成 MFSVM 不能有效对安全等级进行分类。采用网格搜索法对参数进行优化, 将  $C$  和  $\sigma$  的取值在一定范围内划分为网格, 遍历所有网格内的值, 利用 5 折交叉验证法 (cross validation, CV) 来验证每一个  $C$  和  $\sigma$ , 最终选取具有最佳分类精度的  $C=1$ ,  $\sigma$  的取值空间为  $\{0.001, 0.01, 0.11, 10, 100\}$ 。

#### 1.3 评价指标

本研究主要针对粮食供应链危害风险预警开展研究, 需要将输入的多维异构数据分类为安全等级、较安全等级、预警等级、较低危险等级、中危险等级、较高

危险等级、高危险等级、超高危险等级 8 个类别。因此实验采用准确率、运行时间以及混淆矩阵 3 个方面来进行预警模型的评价。对于分类任务, 使用真正例 (TP)、真负例 (TN)、假正例 (FP) 和假负例 (FN) 来比较分类器的预测结果与专家评价的结果。其中, 正例和负例指的是分类器结果, 而真和假指的是专家评价的结果, 如 TP 是指分类器预测的等级与专家评价的等级都是安全等级或者别的等级。计算得到的准确率是分类模型所有预测正确的样本数占总预测样本数的比值, 其表达式见式 (14)。

$$\text{准确率}/\% = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100 \quad (14)$$

运行时间是模型训练时间与模型测试时间的总和。准确率越高, 运行时间越短, 说明模型的性能越好。混淆矩阵是衡量分类型模型准确度的直观方法, 可用以评判模型结果的指标好坏。

## 2 结果与分析

### 2.1 不同模型性能对比

将 21 027 个样本当作训练集, 用于训练整个预警模型, 5 257 个样本作为测试集, 用于测试整个预警模型的准确性。实验环境为 Windows 1 064 位系统, 所用处理器为 Intel (R) Core (TM) i5-4210M CPU@2.60GHz, 运行内存为 8 GB, 数据加速器使用 1 块 NVIDIA RTX 2080 Ti。实验程序设计语言为 Python 及相关库, 训练框架为 PyTorch 和 Keras 深度框架。为了验证本模型的准确性, 并将其与传统的机器学习方法 K 最近邻、朴素贝叶斯、逻辑回归、支持向量机、决策树、随机森林以及 DBN 模型和 BP 神经网络进行对比测试, 在训练集上进行训练, 之后在验证集上执行评估以最小化过度拟合, 当实现训练过程和参数的最佳选择时, 在未知的测试集上进行最终评估, 通过分析不同的性能指标, 评估不同算法结果的可靠性。

表 2 DBN-MFSVM 与传统机器学习方法性能对比结果  
Table 2 Comparative performance of DBN-MFSVM and traditional machine learning methods

方法	准确率/%	运行时间/s
K 最近邻	57.86	92
朴素贝叶斯	51.58	87
逻辑回归	81.36	90
支持向量机	84.38	88
决策树	81.64	86
随机森林	85.76	85
BP 神经网络	90.82	93
DBN	92.65	90
DBN-MFSVM	<b>98.44</b>	85

注: 加粗表示该模型与其他模型对比准确率更高。



表2比较了不同方法准确率和运行时间对比结果,可以看出,DBN-MFSVM的准确率达到98.44%,性能优于其他传统机器学习方法,这表明DBN-MFSVM模型面对多维异构的食品抽检数据集,能够自适应处理数据,形成预警相关的高维度特征,并深入挖掘各危害因素在供应链环节中内在关联,避免了手工提取选择特征的繁重工作,提升危害物风险等级预警的准确率。而其他传统方法预警精度较低,且运行时间稍微较长,进一步证明本实验所采取的方法具备快速高效预警特性,可为粮食产品危害物监管实际应用提供参考。

## 2.2 损失函数和误差对比分析

为进一步说明本实验方法在预警应用上的鲁棒性,对比DBN和BP神经网络在数据集上的损失函数变化趋势,以训练过程中的损失函数值为纵坐标,以训练过程中向前迭代次数为横坐标,得到图2。

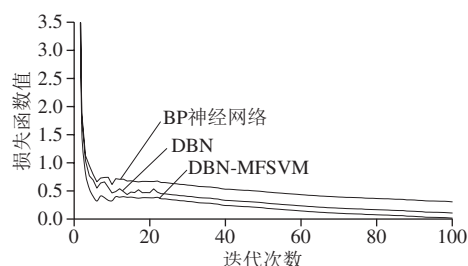


图2 DBN-MFSVM、DBN和BP神经网络的损失函数图

Fig. 2 Loss function graph of DBN-MFSVM, DBN and BP neural network

图2中,损失函数值越小,表明模型的鲁棒性越好,可以看出在迭代训练过程中,3个模型的损失函数值呈下降趋势,均稳定了100次迭代。其中,DBN-MFSVM模型的损失函数值下降最快,最终损失函数下降到0.0162左右,表明该方法具有较强的稳定性能。而变化最慢的是BP神经网络损失函数,只降至0.3062左右,性能远低于本实验模型。为进一步对比说明,从测试数据中随机选取100个样本进行分析,绘制DBN-MFSVM、DBN和BP的误差曲线(图3)。

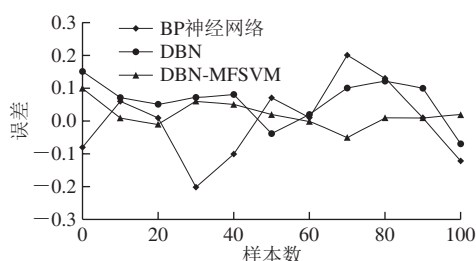


图3 DBN-MFSVM、DBN和BP神经网络的误差曲线图

Fig. 3 Error graph of DBN-MFSVM, DBN and BP neural network

由图3可知,本实验方法的误差范围在0.10~-0.05之间,变化幅度较小,模型对不同样本的分类准确率较高,且整体呈现稳定趋势。而其他方法则存在较大误差波动,

意味对比方法在面对食品数据的分析预警时,容易出现误报错报的行为。

## 2.3 混淆矩阵分析

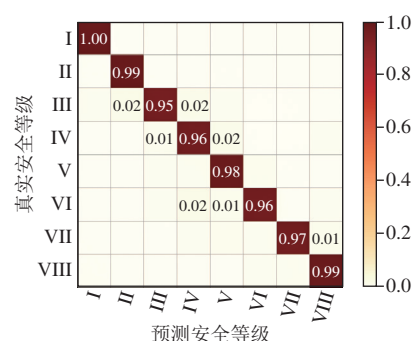


图4 DBN-MFSVM模型的混淆矩阵

Fig. 4 Confusion matrix of DBN-MFSVM model

混淆矩阵反映了模型预警结果的性能,行代表真实的安全等级,列代表模型预测的安全等级,对角线为被正确预测的样本比例。由图4可知,大部分预测正确的样本都集中在对角线附近,说明本实验模型对整个多维异构食品数据的分级预警效果很好,但在预警等级(III)、较低危险等级(IV)和中危险等级(V)中存在一定的混淆偏差,其混淆程度分别为0.95、0.96和0.98,其不正确分类实例多数被分配到较低危险类别(IV)和中危险类别(V),说明本模型在这两个风险级别的存在预警偏差的情况,这就为供应链中重点监控提供了方向,可以起到节省监控成本的作用。

## 2.4 粮食食品供应链主要危害物风险预警系统

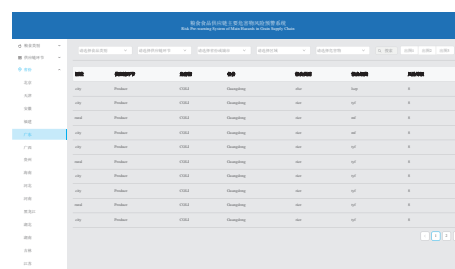


图5 粮食食品供应链主要危害物风险预警系统

Fig. 5 Risk pre-warning system of main hazards in grain supply chain

基于以上的结果,将基于DBN-MFSVM的粮食供应链危害物风险预警方法结合Web技术构建了粮食食品供应链主要危害物风险预警系统(图5)。本系统旨在为广大粮食消费者提供一个可靠的、直观的粮食食品安全信息。本系统采用前后端分离的方式进行开发,其中前端使用React框架,版本为16.9.0;后端使用Django框架,版本为2.2。在本系统选择某些抽检数据时,前端会通过接口向后端发送请求,将选择的数据带到后端,后端通过调用本实验的方法给出预警等级,并将查询出的数据返回给前端,前端将数据转换成所需的数据结构,将其渲

染到表格中, 让用户能清晰地查询到每条数据的情况。为消费者提供更直观快速的粮食食品安全评测结果以及为粮食安全监管机构提供针对性的粮食抽检策略。

### 3 结 论

粮食供应链安全是多维、复杂、耦合的过程, 受到食品多样性、数据多源异构、供应链环节等因素影响, 存在被食源性危害物如重金属、农药残留、微生物、霉菌毒素等风险威胁。现有风险预警受到监管目的性、数据可及性及全局联动性等因素影响, 存在一定应用局限性。本实验在分析全国26个省份的大量抽检数据及关联信息基础上, 建立了基于DBN-MFSVM的风险分级预警模型。首先采用嵌入归一化的数据预处理方法将原始数据处理为结构化数据, 然后输入到DBN模型进行高维度特征提取, 自适应地挖掘供应链中各危害因素间风险变化及内在关联概率, 最后将二分类模糊支持向量机扩展为多类性能, 形成MFSVM分类器, 经过模型训练和参数优化, 实现供应链中各主要危害物风险分级预警。对比结果显示, 本实验模型综合风险预警准确率达到98.44%, 性能优于传统机器学习模型, 且对于粮食抽检数据具有更好鲁棒性和泛化性。建立的此风险预警方法能识别出粮食供应链中危害物风险程度和优先次序, 科学分析我国粮食各供应链环节上各类危害物的风险分级情况, 为监管部门制定有针对性的预警策略、确立优先监管领域和合理分配风险管理措施资源提供科学依据, 也为广大消费者在选择粮食产品时减少食品安全风险隐患提供可行指导。

### 参考文献:

- [1] 张永建. 客观准确认识我国食品安全问题深化食品安全监管与治理[J]. 食品科学技术学报, 2014, 32(3): 1-5. DOI:10.3969/j.issn.2095-6002.2014.03.001.
- [2] KE S, CHENG X Y, ZHANG N, et al. Cadmium contamination of rice from various polluted areas of China and its potential risks to human health[J]. Environmental Monitoring & Assessment, 2015, 187(7): 408. DOI:10.1007/s10661-015-4638-8.
- [3] XIAO G X, HU Y L, LI N, et al. Spatial autocorrelation analysis of monitoring data of heavy metals in rice in China[J]. Food Control, 2018, 89: 32-37. DOI:10.1016/j.foodcont.2018.01.032
- [4] 李强, 刘文, 孙爱兰, 等. 我国加工食品风险分级评价研究[J]. 食品与发酵工业, 2015, 41(9): 220-224. DOI:10.13995/j.cnki.11-1802/ts.201509042.
- [5] HU Y N, CHENG H F, TAO S. The Challenges and solutions for cadmium-contaminated rice in China: a critical review[J]. Environment International, 2016, 92/93: 515-532. DOI:10.1016/j.envint.2016.04.042.
- [6] VAN ASSELT E D, VAN DER FELS-KLERX H J, MARVIN H J P, et al. Overview of food safety hazards in the European dairy supply chain[J]. Comprehensive Reviews in Food Science and Food Safety, 2017, 16(1): 59-75. DOI:10.1111/1541-4337.12245.
- [7] WANG J, YUE H L. Food safety pre-warning system based on data mining for a sustainable food supply chain[J]. Food Control, 2017, 73: 223-229. DOI:10.1016/j.foodcont.2016.09.048.
- [8] WU Y N, CHEN J S. Food safety monitoring and surveillance in China: past, present and future[J]. Food Control, 2018, 90: 429-439. DOI:10.1016/j.foodcont.2018.03.009.
- [9] LIU J, SUN L H, ZHANG J C, et al. Aflatoxin B<sub>1</sub>, zearalenone and deoxynivalenol in feed ingredients and complete feed from central China[J]. Food Additives & Contaminants: Part B, 2016, 9(2): 91-97. DOI:10.1080/19393210.2016.1139003.
- [10] 周少君, 顿中军, 梁骏华, 等. 基于半定量风险评估的食品风险分级方法研究[J]. 中国食品卫生杂志, 2015, 27(5): 576-585. DOI:10.13590/j.cjfh.2015.05.021.
- [11] 程铁军, 冯兰萍. 大数据背景下我国食品安全风险预警因素研究[J]. 科技管理研究, 2018, 38(17): 175-181. DOI:10.3969/j.issn.1000-7695.2018.17.027.
- [12] 周萌. 关于完善我国食品安全风险预警系统的探索[J]. 现代食品, 2019(1): 71-73. DOI:10.16736/j.cnki.cn41-1434/ts.2019.01.025.
- [13] 曹静茹. 基于贝叶斯网络的食品供应链的风险研究[J]. 现代食品, 2017(18): 52-55. DOI:10.16736/j.cnki.cn41-1434/ts.2017.18.016.
- [14] 肖克晶, 左敏, 王星云, 等. 改进的关联规则在食品安全预警上的应用[J]. 食品科学技术学报, 2017, 35(2): 89-94. DOI:10.3969/j.issn.2095-6002.2017.02.014.
- [15] 游清顺, 王建新, 张秀宇, 等. 基于支持向量机的食品安全抽检数据分析方法[J]. 软件工程, 2019, 22(2): 29-31. DOI:10.19644/j.cnki.issn2096-1472.2019.02.009.
- [16] 纪思琪, 吴芳, 李乃祥. 基于决策树的蔬菜病害静态预警模型[J]. 天津农学院学报, 2017, 24(2): 77-80. DOI:10.3969/j.issn.1008-5394.2017.02.019.
- [17] 王星云, 左敏, 肖克晶, 等. 基于BP神经网络的食品安全抽检数据挖掘[J]. 食品科学技术学报, 2016, 34(6): 85-90. DOI:10.3969/j.issn.2095-6002.2016.06.015.
- [18] WANG Jing, YUE Huili, ZHOU Zenan. An improved traceability system for food quality assurance and evaluation based on fuzzy classification and neural network[J]. Food Control, 2017, 79: 363-370. DOI:10.1016/j.foodcont.2017.04.01.
- [19] WANG Xueli, ZHOU Moqin, JIA Jinzhu, et al. A bayesian approach to real-time monitoring and forecasting of Chinese foodborne diseases[J]. International Journal of Environmental Research and Public Health, 2018, 15(8): 1740. DOI:10.3390/ijerph1508174.
- [20] ZHU Changxing, WANG Fenge. Study on risk pre-warning model of China food based on SVM classification[C]// International Conference on E-product E-service and E-entertainment, Henan, 2010. DOI:10.1109/ICEEE.2010.5660142.
- [21] 王霞. 基于BP神经网络的食品安全评估系统的开发[J]. 新余学院学报, 2015, 20(3): 11-14. DOI:10.3969/j.issn.2095-3054.2015.03.004.
- [22] 郑剑, 周豪, 巫丹. 食品安全风险预警领域大数据挖掘的应用[J]. 食品安全导刊, 2017(24): 29. DOI:10.16043/j.cnki.cfs.2017.24.020.
- [23] 李宗亮. 基于大数据挖掘的食品安全风险预警系统研究[D]. 长沙: 湖南大学, 2016: 5-10.
- [24] 徐宁, 步江涛. 食品质量安全风险监测预警工作现状与思考[J]. 食品安全导刊, 2018(26): 25-26. DOI:10.16043/j.cnki.cfs.2018.26.011.
- [25] WANG G M, QIAO J F, BI J, et al. TL-GDBN: Growing deep belief network with transfer learning[J]. IEEE Transactions on Automation Science and Engineering, 2018, 16(2): 874-885. DOI:10.1109/tase.2018.2865663.
- [26] 徐路路, 张德贤. 基于深度信念网络的粮食产量预测[J]. 河南工程学院学报(自然科学版), 2019, 31(1): 77-80. DOI:10.16203/j.cnki.41-1397/n.2019.01.016.
- [27] 郭丹, 路阳, 李建宁, 等. 基于深度信念网络的水稻纹枯病识别方法研究[J]. 农机化研究, 2019, 41(12): 42-45; 85. DOI:10.13427/j.cnki.njyi.2019.12.007.
- [28] 程尚坤. 基于深度学习的储粮害虫检测方法研究[D]. 郑州: 河南工业大学, 2017: 22-27.
- [29] 伏为峰, 杨辉华, 刘振丙, 等. 基于深度信念网络与随机森林的药品鉴别方法[J]. 计算机仿真, 2018, 35(4): 325-330. DOI:10.3969/j.issn.1006-9348.2018.04.070.
- [30] 王小艺, 王珍妮, 孔建磊, 等. 多维数据驱动粮食供应链危害物风险综合评价[J]. 食品科学技术学报, 2019, 37(6): 129-138. DOI:10.3969/j.issn.2095-6002.2019.06.018.