

长短期记忆神经网络在肉制品中铅含量风险预警的应用

陈锂¹, 邹礼华², 孟可欣³, 陈翔², 尹佳¹, 董曼¹, 郭鹏程¹, 文红¹

(1. 湖北省食品质量与安全监督检验研究院, 湖北省食品质量安全检测工程技术研究中心, 湖北武汉 430075)

(2. 武汉理工大学计算机科学与技术学院, 湖北武汉 430070) (3. 武汉理工大学信息工程学院, 湖北武汉 430070)

摘要: 本文基于 2015~2018 年国家肉制品监督抽检的 5 万批次检测数据, 使用长短期记忆神经网络模型建立肉制品中铅含量的风险预警模型。参照国家食品检验标准并结合专家打分, 构建肉制品中铅含量的 6 个食品安全风险等级; 运用 Softmax 和汉宁窗对风险等级数据进行预处理, 使用 Tensorflow 建立三层长短期记忆神经网络的时间序列风险预警模型, 通过 500 轮模型训练实现对肉制品中铅含量的风险趋势预测。结果表明, 长短期记忆神经网络对肉制品中铅含量的预测有较高的准确率, 31 个省份的平均误差为 0.27, 同实际检测风险基本匹配; 模型稳定重现性较好, 运行十次的平均误差为 0.27。此模型可以实现对全国不同地域铅含量的趋势预警, 为日常监督抽检和食品安全风险预警提供技术支撑。

关键词: 风险预警; 长短期记忆神经网络; 食品安全; 铅

文章编号: 1673-9078(2020)08-317-324

DOI: 10.13982/j.mfst.1673-9078.2020.8.0129

Application of Long Short-term Memory Neural Network in Early Warning of Lead Risk in Meat Products

CHEN Li¹, ZOU Li-hua², MENG Ke-xin³, CHEN Xiang², YIN Jia¹, DONG Man¹, GUO Peng-cheng¹, WEN Hong¹

(1. Hubei Provincial Institute for Food Supervision and Test, Hubei Provincial Engineering and Technology Research Center for Food Quality and Safety Test, Wuhan 430075, China)

(2. School of computer science and technology, Wuhan University of Technology, Wuhan 430070, China)

(3. School of information engineering, Wuhan University of Technology, Wuhan 430070, China)

Abstract: In this paper, the detection data was derived from 50,000 batches of meat products sampling inspection by the China Food and Drug Administration, from 2015 to 2018. A Long Short-Term Memory (LSTM) neural network model for lead content in meat products was constructed to analyze the current status of meat product safety and the types of risks. Six levels of food safety risk in lead were established by combining the national food safety standards as a reference with experts scoring. Then the food safety risk data were processed by softmax and hanning window. The Tensorflow was applied to establish a three-layer LSTM neural network of time series risk early warning model through 500 rounds of model training. The results showed that the LSTM had a high accuracy for risk warning in lead content, the data from 31 provinces basically matched the actual risk value, with an average error of 0.27. The model was a stable and reproducible method for forecasting as well, with the average error of ten runs is 0.27, which could provide a new method to implement the trend early-warning in different regions of the country and provide a technical support for government and related authorities for daily supervision.

Key words: long short-term memory neural network; food security; risk warning; lead

引文格式:

陈锂, 邹礼华, 孟可欣, 等. 长短期记忆神经网络在肉制品中铅含量风险预警的应用[J]. 现代食品科技, 2020, 36(8): 317-324

CHEN Li, ZOU Li-hua, MENG Ke-xin, et al. Application of long short-term memory neural network in early warning of lead risk in meat products [J]. Modern Food Science and Technology, 2020, 36(8): 317-324

收稿日期: 2020-02-14

基金项目: 国家重点研发计划项目 (2018YFC1603602)

作者简介: 陈锂 (1988-), 男, 博士, 工程师, 研究方向: 食品安全风险评估与预警

肉制品是指以畜禽肉为主要原料,经过调味制作的熟肉制成品或半成品,是具有营养价值的重要日常消费食品,其食用安全性关系到广大人民群众的健康问题。由于原料污染或生产加工过程中的引入,肉制品中的重金属含量为重要的日常食品监测项目之一^[1],而铅含量是检出率和食品安全风险较高的参数,若长期食用铅污染严重的食品,会损害人体健康并导致疾病发生。随着我国对食品安全的日益重视,食品监督抽检和风险监测力度不断提升。2015-2018年,我国公开了约90.9万批次的食品监督抽检数据,其中肉制品约5万批次,占整体抽检数据的5.5%,大量的监督抽检数据为研究食品安全现状、制定监督抽检计划和提高食品风险防控提供了重要的基础数据,但在食品检测数据的应用方面我国还处于起步阶段。而发达国家和地区基于食品安全监管数据构建了成熟的食源性疾病、食物污染等监测与预警系统,包括世界卫生组织建立的“国际食品安全当局网络”(International Food Safety Authorities Network, INFOSAN)、全球环境监测系统(Global Environmental Monitoring System, GEMS)的食品污染监测与评估规划(GEMS/Food)、美国食品药品监督管理局(Food and Drug Administration, FDA)推行的“危害分析与关键控制点制度”(Hazard Analysis Critical Control Point, HACCP)、欧盟建立的欧盟食品与饲料快速预警系统(Rapid alert system for food and feed, RASFF)等,实践表明基于数据分析的食品安全检测与预警可以有效的对未发生的风险进行提前判断,实现食品安全的源头防控和主动预防^[2-5]。

目前,我国食品安全抽检主要有国家级抽检、省级抽检、国家转移地方性抽检和市县级抽检四个级别,每年可产生几千万条数据记录,而监管机构主要通过传统的人工方式使用数理统计、典型病例通报等手段对食品进行风险警示,对于大量的食品抽检监测数据的深入分析与应用,缺乏有效的数据处理手段^[6]。由于食品检测数据的复杂性和无序性,如何有效的利用数据进行风险预警并拥有一定的精度,一直是食品风险预警研究的重要方向。目前,段文仲^[7]使用模糊数学的理论,建立隶属度级,通过模糊集获得一个评判级,参照评判集来确定食品安全风险等级;刘於勋等运用灰色系统原理获取的外延对象信息进行食品安全风险评价^[8];LIU TongJuan使用神经网络、分类回归

树和贝叶斯网络等算法,通过组合模型建立了易腐烂食品在运输过程中的变质监测预警模型^[9]。人工神经网络(artificial neural network, ANN)计算模型是在现代神经科学研究成果基础上提出的,通过模拟生物神经网络初拉力、记忆信息的方式对信息进行处理,拥有学习和自适应性,已被广泛的应用于许多领域^[10-12]。已有学者开发了食品生产供应链中异常监测数据的神经网络数学模型^[13];用于分子质量预测的食品酶解过程的人工神经网络模型^[14];基于监督抽检数据的食品不合格指标神经网络预测模型等^[15]。因此,基于神经网络的食品安全预警方法,可以有效的在实际中进行应用。

本研究以2015-2018年肉类食品国家公开监督抽检信息为基础,运用人工神经网络中的长短期记忆网络模型(LSTM, Long Short-Term Memory),依据国家标准对检测数据进行风险等级划分,采用数据分箱和数据平滑对风险等级数据进行预处理,建立时间序列的多城市3层LSTM网络风险预测模型,并对模型进行有效验证。重点监测全国肉制品中铅含量并发出风险预警,为我国肉类食品重金属安全提供风险依据和风险预警理论支持,为食品安全大数据的进一步深入挖掘提供参考。

1 材料与方法

1.1 实验材料

本文使用的数据来源于市场监督管理局2015-2018年已对外公开的肉制品监督抽检信息,数据涵盖预制肉制品、熟肉制品等各种产品类型,包括抽样单位、抽样场所、抽样省份、生产许可证、样品名称、抽样编号、食品类别、生产日期、抽样时间、检验项目、检验结果、结果判定、检验依据、判定依据、标准方法、标准检出限、标准方法限、检出限单位、标准最大允许限量等83项原始信息,本文主要以铅检测数据为研究对象,其部分所得数据如表1所示,表中检验结果为实际检测值。

2015-2018年肉与肉制品所有原始数据中均含有铅检测项目,共计约5万条数据,其中铅含量有检测值数据约2.9万条,不合格数据16条,其余为未检出数据。所有数据按省市和生产日期进行划分,每半个月数据作为一个数据集进行下一步处理。

表 1 部分食品安全监督抽检监测原始数据汇总表

Table 1 Summary table of partial raw data of safety supervision and sampling inspection								
省份	生产日期	食品亚类	检验项目	检验结果 (mg/kg)	判定结果	判定依据	标准最大限 量/(mg/kg)	方法检出限 (mg/kg)
福建	2018/4/26	熟肉制品	铅(以 Pb 计)	0.30	合格项	GB2762	0.5	0.04
浙江	2018/7/30	熟肉制品	铅(以 Pb 计)	0.33	合格项	GB2762	0.5	0.04
江苏	2018/6/16	熟肉制品	铅(以 Pb 计)	0.09	合格项	GB2762	0.5	0.04
山东	2018/8/30	熟肉制品	铅(以 Pb 计)	0.31	合格项	GB2762	0.5	0.04
福建	2018/6/14	熟肉制品	铅(以 Pb 计)	0.15	合格项	GB2762	0.5	0.04
贵州	2017/10/30	熟肉制品	铅(以 Pb 计)	0.21	合格项	GB2762	0.5	0.04
安徽	2018/6/25	预制肉制品	铅(以 Pb 计)	0.32	合格项	GB2762	0.5	0.04
山东	2018/7/1	熟肉制品	铅(以 Pb 计)	0.32	合格项	GB2762	0.5	0.04
安徽	2018/6/27	预制肉制品	铅(以 Pb 计)	0.34	合格项	GB2762	0.5	0.04
湖南	2018/6/25	熟肉制品	铅(以 Pb 计)	0.42	合格项	GB2762	0.5	0.04
贵州	2018/2/28	预制肉制品	铅(以 Pb 计)	0.43	合格项	GB2762	0.5	0.04
海南	2018/6/27	熟肉制品	铅(以 Pb 计)	0.04	合格项	GB2762	0.5	0.04
河南	2018/6/12	熟肉制品	铅(以 Pb 计)	0.14	合格项	GB2762	0.5	0.04
北京	2018/4/24	熟肉制品	铅(以 Pb 计)	未检出	合格项	GB2762	0.5	0.04
河南	2018/1/13	熟肉制品	铅(以 Pb 计)	0.47	合格项	GB2762	0.5	0.04
山东	2018/7/16	熟肉制品	铅(以 Pb 计)	0.48	合格项	GB2762	0.5	0.04

1.2 实验方法

1.2.1 长短期记忆网络模型 (LSTM) 工作原理

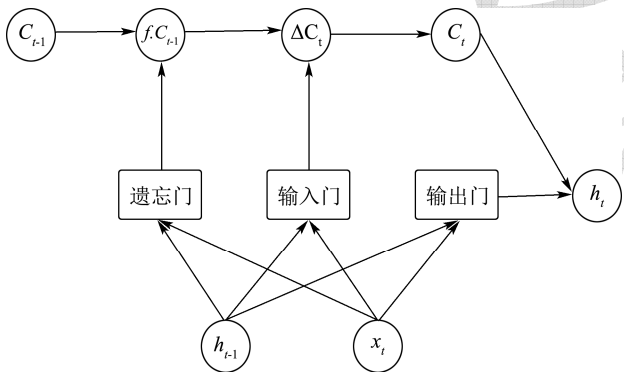


图 1 长短期记忆网络模型计算流程图

Fig.1 Flowchart of LSTM network model

LSTM 长短期记忆模型是对 RNN (Recurrent Neural Network) 循环神经网络模型的一种改进,意在解决随着时间跨度增大时,RNN 模型出现的梯度消失和梯度爆炸等现象,也是一种时间循环神经网络模型。梯度消失可以理解为随着时间跨度的增大,某些神经节点的权重最终会趋向于 0,梯度爆炸则与之相反,随着时间跨度的增大,某些神经节点的权重最终会趋向于无穷大。当这些现象出现时必定会对向后传播过程中的权重更新造成极大的影响,这也是 RNN 循环神经网络模型对长时间跨度序列预测不准的原因之一^[16]。LSTM 对于所有的信息都是有选择的,与传统的

神经网络模型的全盘接受不同,LSTM 会选择遗忘原有信息中不重要的部分,同时吸收当前信息中的重要部分。而为了实现这一点,LSTM 引用了门的机制,用以筛选信息。门的结构是一个简单的神经网络,其输出是一个向量,向量中的元素都是[0,1]的数,用以衡量对信息的接受程度,0 表示拒绝,1 表示完全接受,示意图见图 1。

在 LSTM 中,采用了三个门分别对状态信息、输入信息、输出信息进行控制,它们分别是遗忘门、输入门和输出门^[17]。本研究将前 n 个月的肉制品中的铅含量数据组成一个序列,输入到 LSTM 网络中进行学习。LSTM 网络在预测时会自动计算前面某月份铅含量对后面的月份铅含量的影响,依据该影响决定是否遗忘,并得到输出值。同时整个模型对状态信息进行实时更新,其更新公式(1)如下:

Ct=f_t×C_{t-1}+i_t×ΔC_t (1)

其中: f_t是遗忘门的值, C_{t-1}是前一时刻的铅含量数据, i_t是输入的铅含量数据门的值, ΔC_t是更新的铅含量数据状态信息。

整个公式表达的含义是:当前时间段的肉制品中铅含量数据信息是在保留部分原有的铅含量数据信息的基础上,更新部分新增的铅含量数据信息。而更新的ΔC_t则是由当前输入的铅含量数据 X_t和上次输出的铅含量数据 h_{t-1}决定, b_c为针对状态 C_t的偏置量:

ΔC_t=tanh (W_c[x_t, h_{t-1}]+b_c) (2)

其结构也与 BP 神经网络类似，最终输出的未来时间段的铅含量数据的预测结果 h_t 是由输出门的值 O_t 和状态信息 C_t 决定：

$$h_t=O_t \cdot \tanh(C_t) \tag{3}$$

通过遗忘门负责决定保留多少上一时刻的单元状态到当前时刻的单元状态；输入门负责决定保留多少当前时刻的输入到当前时刻的单元状态；输出门负责决定当前时刻的单元状态有多少输出，实现肉制品中的铅含量趋势预测分析。

1.2.2 数据预处理

数据本身的分布规律和特点对整个神经网络模型

都有较大的影响，恰当的数据预处理可以使预测结果更加准确，并提升模型的运行效率。不恰当的预处理则会导致结果存在较大的偏差，从而难以在生产实践中得到应用。本文以食品检测项目中的大量往期“检验结果”为训练集，由于食品检测结果的随机性和不统一性^[18]，其均值较小，方差大，缺省数据多，容易受到多种噪声的污染。如果直接将原始的生产数据作为模型的输入，学习曲线将十分复杂，同时结果也缺乏可信度。针对食品检测数据的特征，本文主要采用数据分箱和汉宁窗平滑对数据进行预处理。

表 2 风险等级分类表

Table 2 Risk classification table

风险等级	评价值	预警指标	备注
0 级	未检出	安全	无污染
1 级	$0 < \text{检测值} \leq 0.1 \text{ MRL}$	无警	有少量有害物质
2 级	$0.1 \text{ MRL} < \text{检测值} \leq 0.3 \text{ MRL}$	微警	有害物质在可接受范围内
3 级	$0.3 \text{ MRL} < \text{检测值} \leq 0.7 \text{ MRL}$	轻警	有害物质含量较高
4 级	$0.7 \text{ MRL} < \text{检测值} \leq 1.0 \text{ MRL}$	中警	有一定程度危害程度健康
5 级	$1.0 \text{ MRL} < \text{检测值}$	重警	不合格项目严重危害身体健康

注：MRL：国家标准 GB 2762-2017《食品安全国家标准食品中污染物限量》中肉及肉制品对应的最大允许限量。

表 3 肉与肉制品风险等级表

Table 3 Meat and meat products risk scale

省份	生产日期	食品亚类	检验项目	检验结果 (mg/kg)	标准最大限 量/(mg/kg)	方法检出限 (mg/kg)	风险等级
福建	2018/4/26	熟肉制品	铅(以 Pb 计)	0.30	0.5	0.04	3
浙江	2018/7/30	熟肉制品	铅(以 Pb 计)	0.33	0.5	0.04	3
江苏	2018/6/16	熟肉制品	铅(以 Pb 计)	0.09	0.5	0.04	2
山东	2018/8/30	熟肉制品	铅(以 Pb 计)	0.31	0.5	0.04	3
福建	2018/6/14	熟肉制品	铅(以 Pb 计)	0.15	0.5	0.04	2
贵州	2017/10/30	熟肉制品	铅(以 Pb 计)	0.21	0.5	0.04	3
安徽	2018/6/25	预制肉制品	铅(以 Pb 计)	0.32	0.5	0.04	3
山东	2018/7/1	熟肉制品	铅(以 Pb 计)	0.32	0.5	0.04	3
安徽	2018/6/27	预制肉制品	铅(以 Pb 计)	0.34	0.5	0.04	3
湖南	2018/6/25	熟肉制品	铅(以 Pb 计)	0.42	0.5	0.04	4
贵州	2018/2/28	预制肉制品	铅(以 Pb 计)	0.43	0.5	0.04	4
海南	2018/6/27	熟肉制品	铅(以 Pb 计)	0.04	0.5	0.04	1
河南	2018/6/12	熟肉制品	铅(以 Pb 计)	0.14	0.5	0.04	2
北京	2018/4/24	熟肉制品	铅(以 Pb 计)	未检出	0.5	0.04	0
河南	2018/1/13	熟肉制品	铅(以 Pb 计)	0.47	0.5	0.04	4
山东	2018/7/16	熟肉制品	铅(以 Pb 计)	0.48	0.5	0.04	4

数据分箱是数据预处理中的不可或缺的一环，对连续的数据进行分箱使其离散化，不仅能够极大降低模型的复杂度，同时可以增强对异常值的鲁棒性。本研究根据国家标准 GB 2762-2017《食品安全国家标准

食品中污染物限量》和 GB 5009.12-2017《食品安全国家标准食品中铅的测定》，结合专家打分，对有最大限量的单一项目建立风险指标体系实现对检测数据的分箱，将实际的检测值对应到合适的风险等级并进行预

警。检测值越高风险等级越高,将风险等级分为六级,0级为无污染无风险,1级为无警,2级为轻微风险,3级为轻度风险等级,4级为中度风险等级,5级为不合格检测属于最高风险,等级详见表2。

肉与肉制品中的铅含量是非季节性项目,主要由污染的环境中摄取后通过食物链进行放大产生或在生产、加工、运算和存储环节引入,选取2015-2018年5万条实际检测数据作为实验数据。按照上述风险等级分类表将实际检测数据标准化,见表3。如北京2018年4月24日的熟肉制品铅含量检验结果为未检出,属于0风险产品;福建省2018年4月26日生产的熟肉制品中铅含量检验结果为0.3 mg/kg,大于0.3倍最大允许限量但小于0.7倍,属于第3等级风险;其他产品均以此类推给予风险等级。

通过上述步骤将所有检验数据转换为风险等级,得到5级风险数据16条,4级风险数据约8000条,3级数据约4000条,2级风险数据约1.2万条,1级风险数据约5000条,0级无风险数据约2.1万条。由于

采样的不规律性和数据缺失无法按照每天的数据进行建模,如果以一个月数据进行建模由于数据的不规律性使模型精度较低,因此对数据进行分箱处理,将每半个月数据划分为一个数据集 $X=\{x_i\}$, $x_i=\{x_{ij}\}$ 。使用改进Softmax(公式4)对每个分箱数据集的风险等级进行概率分配,合理的选出对应时间段的代表值。

改进的Soft max =
$$\frac{p * e^{level}}{\sum p * e^{level}}$$
 (4)

其中: p 为铅每个风险等级出现的频率, e 为自然指数的底, $level$ 为等级值。

任意取半个月的铅含量作为一个风险等级数据集,通过改进Softmax公式可得到6个风险等级出现的概率见表4,其中风险等级3在下个预测时间段的概率最大,为4.46345265,故下个时间段的风险等级取值为3。该方法可以有效的提高时间序列模型的预测精度,避免了直接依照概率取值的缺陷,结合风险的危害性对数据建模,更符合实际。

表4 Softmax 分析铅含量风险等级值
Table 4 Lead risk level analyzed by Softmax

风险等级	零级	一级	二级	三级	四级	五级
出现频率/p	0.16666667	0.27777778	0.27777778	0.22222222	0.05555556	0
出现概率	0.16666667	0.75507829	2.05251558	4.46345265	3.03323056	0
预测风险等级	三级					

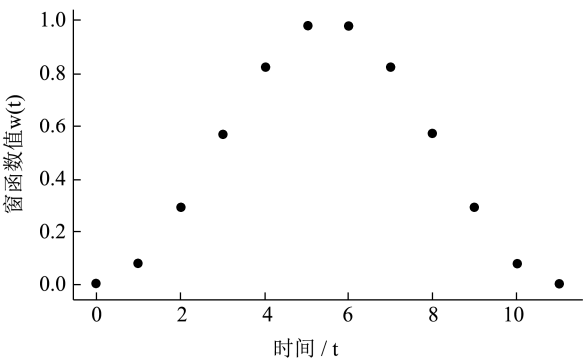


图2 汉宁窗函数示意图

Fig.2 Diagram of Hanning window function

将表4提取的风险等级进行建模,由于数据的不规律性导致模型精度较差、曲线突兀、学习效果不好。为提高模型精度,本研究加入汉宁窗对食品等级序列集进行数据平滑,通过去噪来降低学习的难度。汉宁窗的本质是一种人为的近似处理,属于窗函数的一种,常用于数字信号处理和声波处理中,具有降噪和平滑信号的功能,但存在数据失真的缺点^[19]。因此,窗口的大小与平滑程度呈正相关,窗口越小,平滑程度越低,数据失真越小。从理论上讲,选择窗口大小的原则是越小越好。其本质是一个单位权重的加权函数,

T 为函数的周期,时域表达式见公式5,汉宁窗示意图见图2。

$$w(t) = \frac{1}{2} \left[1 + \cos \left(2\pi \frac{t}{T} \right) \right]$$
 (5)

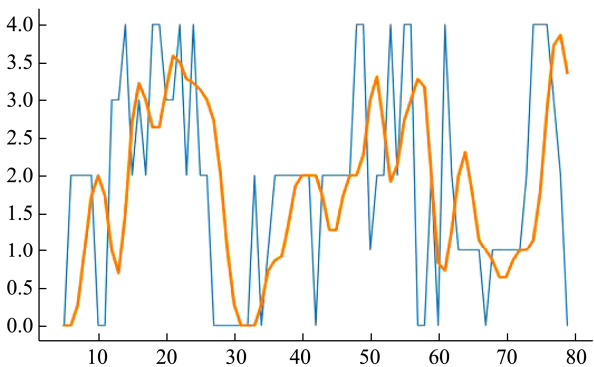


图3 数据平滑示意图

Fig.3 Data smoothing diagram

图3为风险等级数据平滑示意图,蓝线为未平滑的原始风险等级数据,从图中可以看出数据波动较大,无明显规律。黄线为使用汉宁窗平滑后的数据,其曲线较平滑可有效降低模型学习难度,且走势基本同原始数据相吻合。

1.3 模型构建

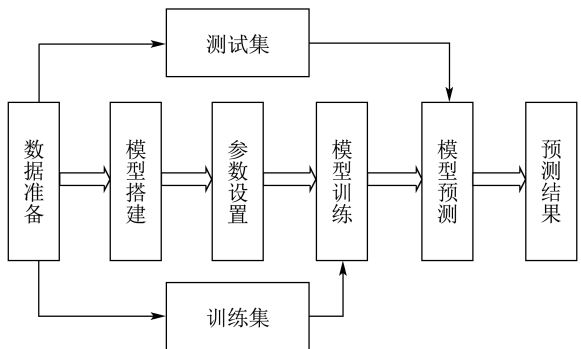


图4 模型构建流程图

Fig.4 Flowchart of model construction

本研究选择使用 Tensorflow 作为 LSTM 模型的搭建平台，以 2015 年 1 月至 2018 年 12 月 32 个省肉制品的铅含量数据作为数据集，将数据集的前三分之二作为训练集，后三分之一为测试集。具体的流程为：数据准备、网络层搭建、参数配置、模型训练和模型预测等 6 个步骤，见图 4。

数据准备阶段，为方便模型的训练并防止过拟合，通过 MinMax 归一化方法（公式 6）将数据映射到[0,1]区间。将一维的数据作张量维度转换为三维的张量数据，作为模型的输入端。其中第一维是样本的数量，第二维是特征量的数量，第三维是输出值的维数。需要说明的是 LSTM 的特征量是预测下一个值所需要的数据量个数，名为 look_back。例如，设置 look_back=6，表明需要使用 6 个连续的数据来预测第 7 个数据，即使用第一个数据到第六个数据来预测第七个数据，或从第二个数据到第七个数据预测第八个数据，以此类推。

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (6)$$

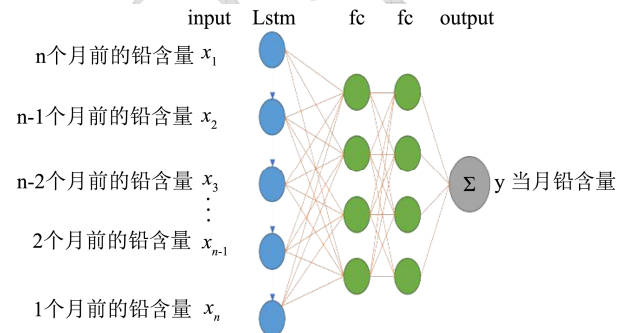


图5 肉制品中铅含量的长短期记忆网络模型的网络结构

Fig.5 LSTM network model for lead content in meat products

注：fc: full connection.

网络层搭建阶段，使用 Tensorflow 中输出为 32 个神经元的 LSTM 层作为模型的第一层，全连接着含

有 32 个神经元的第二层和 16 个神经元的第三层，最后一个单神经元的全连接层作为输出层，见图 5。

参数配置阶段，使用均方误差 (mean-square error, MSE) 作为损失函数，Adam 优化算子^[20]作为向后传播更新参数的算子，将 Adam 优化器的学习率参数设定为 0.001，总共训练 500 轮，学习率在训练期每 100 轮折半减小。为提高模型训练精度将学习设置为动态减小的模式，使用训练集对模型进行训练，测试集检测模型的预测准确度。

2 结果分析

2.1 LSTM 模型有效性分析

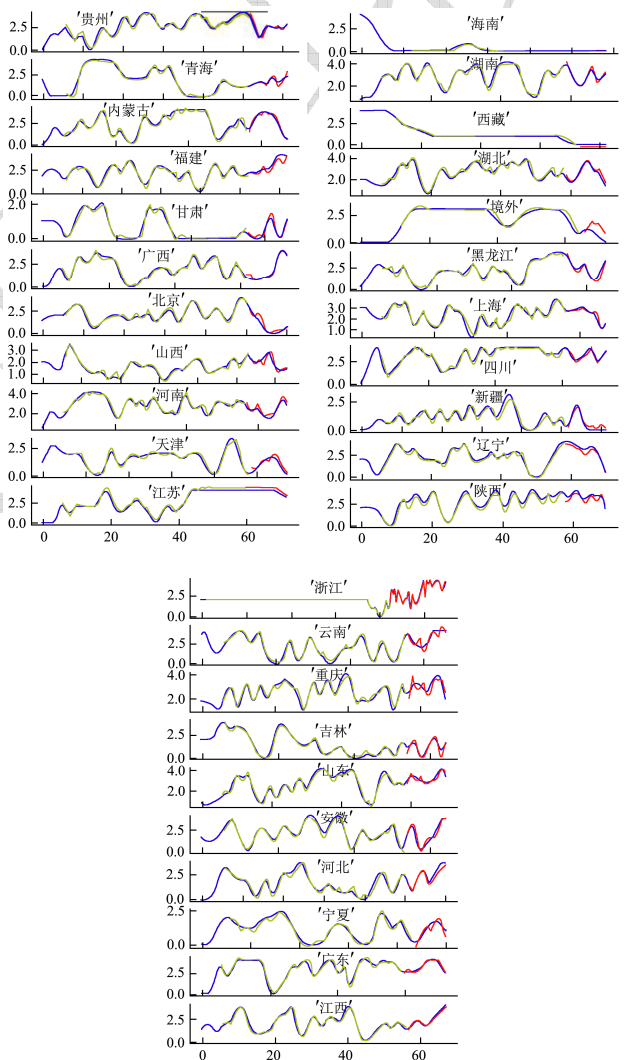


图6 32 个区域肉制品铅含量预测模型图

Fig.6 Prediction model of lead content in meat products in 32 districts

将 2015 年 1 月至 2018 年 12 月份 32 个不同地域肉制品的铅含量数据带入 LSTM 模型进行预测，可得到全国 31 个省份和境外的铅含量预测值（图 6）。蓝

线为原始数据趋势图,黄线为模型训练趋势图,红线为模型预警趋势图。从图中可以看出预测风险等级与实际风险等级匹配较好,精确度较高。使用均方差法计算实际风险等级和预测风险等级的误差值,铅含量风险等级整体误差为 0.27,境外误差较大为 0.65,海南误差最小为 0.07,其余省份误差均低于 0.5。其中,境外误差较大原因为数据量较小,前后规律性不强且差别较大,模型学习不足,从而影响了模型精度,但其它 31 个省份模型学习能力较强,预测等级精确,表明此模型适合于食品中铅含量的时间序列研究。因此,基于长短期记忆神经网络模型的食品中铅风险等级预测有效,可以实现食品中不同省份铅含量的风险等级

趋势预警。

2.2 LSTM 模型稳定性分析

人工神经网络的是食品安全预警研究中的常见算法,由于其初始权值的随机分配性,模型预测值可能出现一定的误差。本研究以 2015-2018 年检测数据前三分之二为模型学习数据,后三分之一为预测风险等级,将预测风险等级同实际风险等级对比,连续重复运行 LSTM 模型 10 次,从表 5 中可以得出 32 个区域铅风险等级平均误差为 0.27 低于 0.5,准确率较高,模型波动较小,此模型运行结果稳定可靠。

表 5 LSTM 模型 10 次运行结果

Table 5 Results of 10 runs of LSTM model

实验次数	1	2	3	4	5	6	7	8	9	10	平均误差
误差值	0.262	0.272	0.276	0.267	0.262	0.277	0.277	0.266	0.267	0.266	0.269

2.3 模型比较分析

贝叶斯网络在食品预警中也有较多的应用,其主要通过结构网络和条件概率表来得到一个分类器,通过分类器来实现预警的功能。但是对于规律性较差的连续时间序列数据,预测性能不太理想。而 LSTM 在时间序列模型中,通过其高度的非线性学习可以较好的解决大量检测数据的无规律性,从而提高模型的精度。选取 2015-2018 年 4 个省份的全部风险等级数据放入模型中进行训练,连续预测 2018 年 11 月和 12

月份的肉制品中铅含量的风险等级。从表 6 中可以得出,LSTM 模型预测的风险等级和实际检测值转换的风险等级比较接近,11 月份平均误差为 0.13,12 月份平均误差为 0.18。贝叶斯网络模型对于有规律的连续型数据可以得到较好的结果,但是当数据出现较大波动时其预测结果会出现较大误差,其中 11 月份预测平均误差为 0.5,12 月份误差较大平均误差为 1.25。因此,相对于贝叶斯网络模型而言,LSTM 模型更适合于时间序列的食品风险预测。

表 6 LSTM 和贝叶斯网络模型的预测结果

Table 6 Prediction results of LSTM and Bayesian network models

风险等级	江西 11 月	江苏 11 月	浙江 11 月	陕西 11 月	江西 12 月	江苏 12 月	浙江 12 月	陕西 12 月
实际值	3	3	3	4	3	3	4	3
贝叶斯网络	3	2	2	4	2	2	2	4
LSTM	2.9	3.2	3.1	3.9	2.9	3.1	4.3	2.8

3 结论

本文基于长短期记忆神经网络模型,结合日常监管工作中大量的肉制品检测数据,通过国家食品安全标准给定的最大限量结合专家打分,首次将检测数据转化为食品安全风险等级作为预测模型的输入和输出,将传统模型预测结果的合格或不合格,提高到 6 个不同的食品安全风险等级。联合使用了改进型 Softmax 和汉宁窗对数据进行分箱和平滑等预处理,有效的解决了食品检测数据的随机性和不统一性带来的模型误差。构建了 31 个省份肉制品中的铅含量时间风险预警模型,模型精度和稳定性较好,可以有效的实现肉制品中铅含量的趋势预测,对大量的检测数据进行了有

效的应用,为我国肉与肉制品的日常风险监管工作提供了有效的数据支持,对完善食品安全预警提供了新的技术方法。进一步的工作是结合并优化数据预处理方法,构建多元数据多维度预测模型,根据数据的变动及时的调整模型网络结构,搭建不同食品的风险预警体系。

参考文献

- [1] 朱有涛,张遐,邵梅,等.微波消解 ICP-MS 法检测肉制品中的铅铬镉铝锰锡[J].食品研究与开发,2018,39(14):172-176
- ZHU You-tao, ZHANG Xia, SHAO Mei, et al. Determination of plumbum, chromium, cadmium,

- manganese, aluminum and tin in meat product by microwave digestion with ICP-MS method [J]. Food Research and Development, 2018, 39(14): 172-176
- [2] Savelli C J, Bradshaw A, Ben Embarek P, et al. The FAO/WHO international food safety authorities network in review, 2004-2018: learning from the past and looking to the future [J]. Foodborne Pathogens and Disease, 2019, 16(7): 480-488
- [3] 付文丽,陶婉亭,李宁,等.借鉴国际经验完善我国食品安全风险监测制度的探讨[J].中国食品卫生杂志,2015,27(3): 271-276
- FU Wen-li, TAO Wan-ting, LI Ning, et al. Improve the food safety risk monitoring system of China from international experience [J]. Chinese Journal of Food Hygiene, 2015, 27(3): 271-276
- [4] Dzwolak W. Assessment of HACCP plans in standardized food safety management systems - the case of small-sized polish food businesses [J]. Food Control, 2019, 106: 106716
- [5] 李笑曼,臧明伍,赵洪静,等.基于监督抽检数据的肉类食品安全风险分析及预测[J].肉类研究,2019,33(1):42-49
- LI Xiao-man, ZANG Ming-wu, ZHAO Hong-jing, et al. Analysis and prediction of meat product safety based on supervision and sampling data [J]. Meat Research, 2019, 33(1): 42-49
- [6] 韩娟,卢林纲,秦玉昌.我国食品和农产品质量安全风险评估体系研究和构建概况[J].食品安全质量检测学报,2016, 7(4):1720-1724
- HAN Juan, LU Lin-gang, QIN Yu-chang. Investigation and construction of food and agricultural products' quality and safety risk assessment system [J]. Journal of Food Safety and Quality, 2016, 7(4): 1720-1724
- [7] 段文仲.模糊综合评价在检验检疫风险管理中的应用[J].检验检疫科学,2006,16(1):16-18
- DUAN Wen-zhong. Application of fuzzy comprehensive evaluation in inspection and quarantine risk management [J]. Inspection and Quarantine Science, 2006, 16(1): 16-18
- [8] 刘於勋.食品安全综合评价指标体系的灰关联分析[J].华北水利水电学院学报,2007,28(4):100-103
- LIU Wu-Xun. Grey correlation analysis of comprehensive food safety evaluation index system [J]. Journal of North China Institute of Water Conservancy and Hydroelectric Power, 2007, 28(4): 100-103
- [9] LIU Tongjuan, HU Anqi. Model of combined transport of perishable food stuffs and safety inspection based on data mining [J]. Food and Nutrition Sciences, 2017, 8(7): 760-777
- [10] 蔡强,王君君,李海生,等.基于神经网络的食物安全评价模型构建研究[J].食品科学技术学报,2014,32(1):69-77
- CAI Qiang, WANG Jun-jun, LI Hai-sheng, et al. Research on establishment of food safety evaluation model based on neural network [J]. Journal of Food Science and Technology, 2014, 32(1): 69-77
- [11] Geng Z, Shang D, Han Y, et al. Early warning modeling and analysis based on a deep radial basis function neural network integrating an analytic hierarchy process: a case study for food safety [J]. Food Control, 2019, 96: 329-342
- [12] Abdel-Nasser M, Mahmoud K. Accurate photovoltaic power forecasting models using deep LSTM-RNN [J]. Neural Computing and Applications, 2019, 31(7): 2727-2740
- [13] Zhang K, Chai Y, Yang S X, et al. Pre-warning analysis and application in traceability systems for food production supply chains [J]. Expert Systems with Applications, 2011, 38(3): 2500-2507
- [14] 续晓琪,薛长湖,常耀光.基于人工神经网络的海参岩藻聚糖硫酸酯酶解模型建立及抗氧化活性研究[J].中国食品学报,2019,19(1):20-27
- XU Xiao-qi, XUE Chang-hu, CHANG Yao-guang. Artificial neural network-based establishment of sea cucumber fucoidan sulfate sulfatase model and its antioxidant activity [J]. Journal of Chinese Institute of Food Science and Technology, 2019, 19(1): 20-27
- [15] Goni S M, Oddone S, Segura J A, et al. Prediction of foods freezing and thawing times: artificial neural networks and geneticalgorithm approach [J]. Journal of Food Engineering, 2008, 84(1): 164-178
- [16] Yu Y, Si X, Hu C, et al. A review of recurrent neural networks: LSTM cells and network architectures [J]. Neural Computation, 2019, 31(7): 1235-1270
- [17] Zhang Y. Early warning method of tariff recovery risk based on long short-term memory neural network [C]//2019 International Conference on Smart Grid and Electrical Automation (ICSGEA). IEEE, 2019: 433-436