

基于决策树排序的挂面中潜在危害物质风险分析

生吉萍^{1*}, 李苗苗¹, 肖革新²

(1. 中国人民大学农业与农村发展学院, 北京 100872; 2. 国家食品安全风险评估中心, 北京 100022)

摘要: 目的 对挂面中潜在危害物质风险进行分析, 对最相关的危害物质进行危害程度排序。**方法** 基于全国 2015~2017 年的挂面危害物检测数据, 利用决策树法将挂面潜在危害物质进行排序。建立了 3 个决策树, 分别对污染物、食品添加剂和生物毒素进行排序。将监测物质分为高、中、低优先级, 以利于在风险评估时确定优先考虑的次序。**结果** 根据对挂面抽检结果的风险分析, 铅、铝、二氧化钛和乙二胺四乙酸二钠、黄曲霉毒素 B₁ 和脱氧雪腐镰刀菌烯醇为高优先级检测物质, 而柠檬黄、日落黄、诱惑红和亮蓝、苋菜红、胭脂红为中优先级检测物质。**结论** 决策树方法有助于对挂面中潜在危害物质的风险优先级排序, 可以作为建立基于风险的监测计划的工具。该方法具有可拓展性, 可以利用最新监测数据, 对其他类型粮食的风险等级进行评估, 得到其潜在危害物质的优先次序物质清单。

关键词: 食品安全危害监测; 决策树; 挂面; 潜在危害物质; 风险分级

Risk analysis of potentially hazardous substances in vermicelli based on decision tree sorting

SHENG Ji-Ping^{1*}, LI Miao-Miao¹, XIAO Ge-Xin²

(1. School of Agricultural Economics and Rural Development, Renmin University of China, Beijing 100872, China;
2. China National Center for Food Safety Risk Assessment, Beijing 100022, China)

ABSTRACT: Objective To analyze the risk of potential hazardous substances in vermicelli and rank the most relevant hazardous substances. **Methods** Based on the nationwide detection data of vermicelli hazard from 2015 to 2017, the potential hazard substances in vermicelli were sorted by the decision tree method. Three decision trees were established to rank pollutants, food additives, and biotoxins respectively. The monitoring substances were divided into high, medium and low priority, so as to determine the priority in risk assessment. **Results** According to the risk analysis of the results of the random inspection of the noodles, lead, aluminum, titanium dioxide, disodium EDTA, aflatoxin B₁ and deoxynivalenol were the high priority detection substances, while lemon yellow, sunset yellow, alluring red and bright blue, amaranth and carmine were the medium priority detection substances. **Conclusion** The decision tree method is helpful to rank the risk priority of potential hazardous substances in vermicelli, and can be used as a tool to establish risk-based monitoring plan. The method is extensible, and can use the latest monitoring data to assess the risk levels of other types of food to obtain a list of priority substances for potential hazardous substances.

KEY WORDS: food safety hazard monitoring; decision tree; vermicelli; potentially hazardous substance; risk

基金项目: 国家重点研发计划项目(2018YFC1603300)、国家自然科学基金重点项目(71633005)

Fund: Supported by National Key Research and Development Plan Project(2018YFC1603300), and National Natural Science Fund Key Project (71633005)

*通讯作者: 生吉萍, 教授, 主要研究方向为食品安全与食品经济管理研究。E-mail: shengjiping@126.com

*Corresponding author: Sheng Ji-Ping, Professor. School of Agricultural Economics and Rural Development, Renmin University of China, Beijing 100872, China. E-mail: shengjiping@126.com

classification

1 引言

世界各地食品安全问题频发, 不仅给消费者的健康造成损失, 也重创食品业。解决食品安全问题已经成为世界各国政府的重要工作内容^[1]。风险分析(risk analysis)是目前国际上公认的控制食品中各种化学性、物理性和生物性危害和突发事件应该遵循的框架, 应用广泛。风险分析由风险评估(risk assessment)、风险管理(risk management)和风险信息交流(risk communication)3个相互关联的部分组成^[2], 其中, 风险评估是风险分析的核心。《实施卫生与植物卫生措施协议》(Agreement on the Application of Sanitary and Phytosanitary Measures, 简称 SPS 协议)规定, 进口国根据可能采用的 SPS 措施, 对其领土上某些害虫或疾病的进入、存在或传播的可能性, 以及对潜在的生物学和经济影响进行评价, 或对食品、饮料和饲料中的添加剂、污染物、毒素或致病菌的存在对人体和动物的健康可能造成的不良作用进行评估^[3]。然而, 我国食品安全风险监测体系建设和风险评估工作依然处于基础阶段^[4]。

风险分级作为一种初步评估的手段, 可以帮助风险管理者通过比较不同来源风险的相对大小, 确定需优先监管的领域, 并得出需进一步进行经典评估的食品危害^[5]。联合国粮农组织(Food and Agriculture Organization of the United Nations, FAO)和世界卫生组织(World Health Organization, WHO)制定的风险管理一般框架中包含了风险分级, 并且将其作为风险监测和评估的前提之一^[6]。食品风险分级作为风险评估的一种形式, 指综合分析食品的污染水平、膳食暴露情况、对公众健康危害等因素, 对食源性危害风险进行分级排序, 目的在于采取科学的风险分析框架, 在众多复杂的食品安全问题中识别风险的优先次序, 合理分配风险管理措施的有限资源。

决策树算法是数据挖掘分类算法中常见的一种方法。它以树状结构表现, 叶子结点代表一个结论, 内部结点描述一个属性, 从上到下的一条路径, 确定一条分类规则结构简单直观, 容易理解, 有较高的分类精度, 在数据挖掘、机器学习、人工智能等领域等都有广泛的应用^[7]。杨娜^[8]曾利用决策树设计了云南粮食安全预警系统, 得出影响粮食安全的各因素及其对粮食安全等级的影响度排序。基于决策树开发的针对蔬菜静态病害进行风险等级分析的模型, 能够达到较高的黄瓜确诊率^[9]。将决策树应用于药品 GSP 预警管理中, 从药品进、存、销三个环节日常管理中产生的大量数据中挖掘有效信息, 以进行及时预警信息定位和管理^[10]。Fels-Klerx 等^[11]分析了 279 个食品和/或营养危害风险排名, 共有 11 种风险排名方法。该分析表明, 定量风

险排序方法需要对单个化合物进行全面的分析, 耗时长、预算大。由于监测程序包含广泛的残留物和污染物, 这些定量方法不太适合风险排序, 在有限的时间和预算范围内对一系列化合物进行排名最合适的方法之一是流程图或者决策树。挂面是我国居民最主要的面食之一, 对其安全进行监控对于保障市民食品安全具有重要的意义。本研究以挂面为研究对象, 利用决策树排序建立一种透明的、实体化的风险排序方法, 可用于完善符合实际发展的食品安全风险评估机制, 促进食品安全风险分析的顺利开展。

2 材料和方法

2.1 数据来源

本文所用的数据为国家市场监督管理总局官方网站获得的全国 2015~2017 年的挂面危害物检测数据。

2.2 污染物分类方法

基于 GB 2760-2014《食品安全国家标准 食品添加剂使用标准》^[12]、GB 2761-2017《食品安全国家标准 食品中真菌毒素限量》^[13]和 GB 2762-2017《食品安全国家标准 食品中污染物限量》^[14], 参考我国粮食风险监测数据, 确定挂面的检测项目主要包括铅、二氧化钛、乙二胺四乙酸二钠、脱氧雪腐镰刀菌稀醇、铝的残留物、黄曲霉毒素 B₁、柠檬黄、苋菜红、胭脂红、日落黄、诱惑红、亮蓝。按照国标中的规定, 将潜在危害物分为 3 组, 即污染物、食品添加剂和真菌毒素。

2.3 决策树设计方法

决策树包含一个根节点、若干个内部节点和若干个叶节点; 叶节点对应决策结果, 即: 低优先级、中优先级、高优先级。其它每个节点对应一个属性测试。决策树的生成是一个递归算法, 关键在于如何选择最优属性划分。随着划分不断进行, 决策树的分支节点所包含的样本尽可能属于同一类别, 用信息熵来度量样本集合纯度, 假定当前样本集合 D 中第 k 类样本所占比例为 $p_k(k=1, 2, \dots, |y|)$ 。信息熵定义为:

$$Ent(D) = - \sum_{k=1}^{|y|} p_k \log_2 p_k$$

$Ent(D)$ 的值越小, D 的纯度越大。在选定属性划分节点时, 信息增益越大, 代表用该属性所获得的纯度提升越大。信息增益的定义为:

$$Gain(D, a) = Ent(D) - \sum_{v=1}^{|V|} \frac{|D^v|}{|D|} Ent(D^v)$$

其中, a 代表属性, V 代表 a 所有可能的取值, D^v 代表在属性

a 上取值为 a^v 的样本。

对 3 组潜在危害物质中的每一组分别进行决策树的设计和优化, 绘制 3 种独立的决策树, 将挂面中潜在危害物质分为高、中、低 3 个不同等级, 结合决策树的基本原理, 将高、中、低 3 个不同等级设置为叶节点来进行研究。

3 结果与分析

3.1 污染物

GB 2762-2017 中规定无论是否制定污染物限量, 食品生产和加工者均应采取控制措施, 使食品中污染物的含量达到最低水平。本数据中检测项目的污染物有铅。

采用决策树法对污染物进行研究, 从数据中提取出 4 种属性, 分别是:

- (1) 污染物在 GB 2762-2017 中是否有使用限制(标为属性 1)
- (2) 2015 至 2017 上半年检测数据中是否发现该物质不合格数据(标为属性 2)
- (3) 污染物的残留是否会对人体健康造成严重和/或不可逆的不良健康影响

可逆的影响(基因毒性致癌、致畸和激素干扰物质的情况也是如此)(标为属性 3)

(4) 这种物质是否有可能出现在生产挂面的生产系统中(标为属性 4)

这 4 种属性都是离散属性, 均有“是”和“否”2 种属性取值, 根据这些属性来判断优先级的程度。根据现有数据计算出根节点的信息熵为 0.875, 不同属性的信息增益结果见表 1。

表 1 污染物决策树信息增益计算结果
 Table 1 Calculation results of pollutant decision tree information gain

属性	信息增益
1	0.302
2	0.103
3	0.220
4	0.105

属性 1 的信息增益最大, 因此选定属性 1 为根节点。依次计算信息增益, 构建叶子节点, 最终的决策树如图 1。

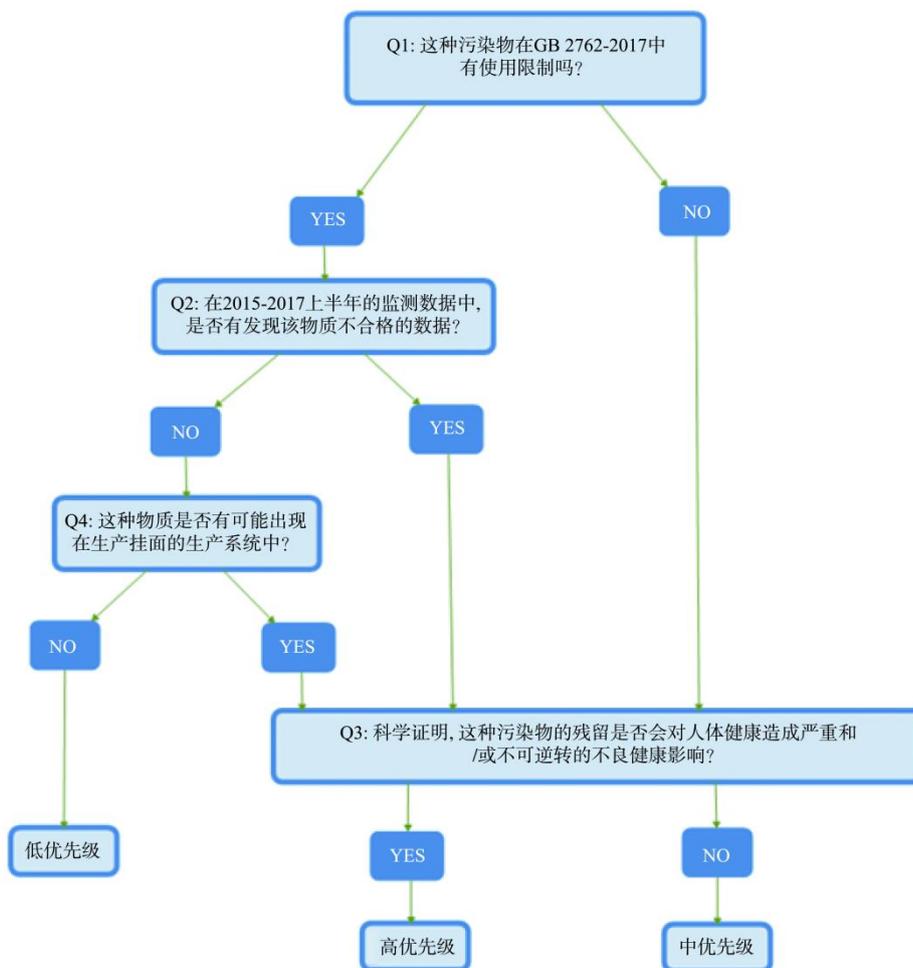


图 1 污染物决策树示意图

Fig.1 Pollutant decision tree

挂面中铅是重要的风险因子。铅在 GB 2762-2017 中有使用限制，且在 2015~2017 年上半年的监测数据中出现该物质不合格的情况。实际生产中，铅常出现在挂面的生产器械中，且科学证明其残留会对人体健康造成严重不良健康影响，故为高优先级。

3.2 食品添加剂

GB 2760-2014 中规定了食品添加剂的使用原则、允许使用的食品添加剂品种、使用范围及最大使用量或残留量。本数据中检测项目的添加剂包括铝、二氧化钛、乙二胺四乙酸二钠、柠檬黄、苋菜红、胭脂红、日落黄、诱惑红、亮蓝。对食品添加剂进行研究时，从数据中提取出 3 种属性，还根据文献增加一个属性，用于研究的 4 个属性分别为：

(1)GB 2760-2014 中对挂面有无这种食品添加剂的限量(标为属性 1)

(2)这种食品添加剂在 GB 2760-2014 中是否有使用限制(标为属性 2)

(3)2015~2017 上半年的监测数据中有没有发现该物质任何不符合国标规定的的数据(标为属性 3)

(4)文献中是否有显示这种食品添加剂的残留会对人体健康造成严重和/或不可逆转的不良健康影响(标为属性 4)

根据现有数据计算出根节点的信息熵为 0.675，不同属性的信息增益结果见表 2。

表 2 食品添加剂决策树信息增益计算结果
Table 2 Calculation results of food additive decision tree information gain

属性	信息增益
1	0.102
2	0.411
3	0.320
4	0.141

属性 2 的信息增益最大，因此选定属性 2 为根节点。依次计算信息增益，构建叶子节点，最终的决策树如图 2。

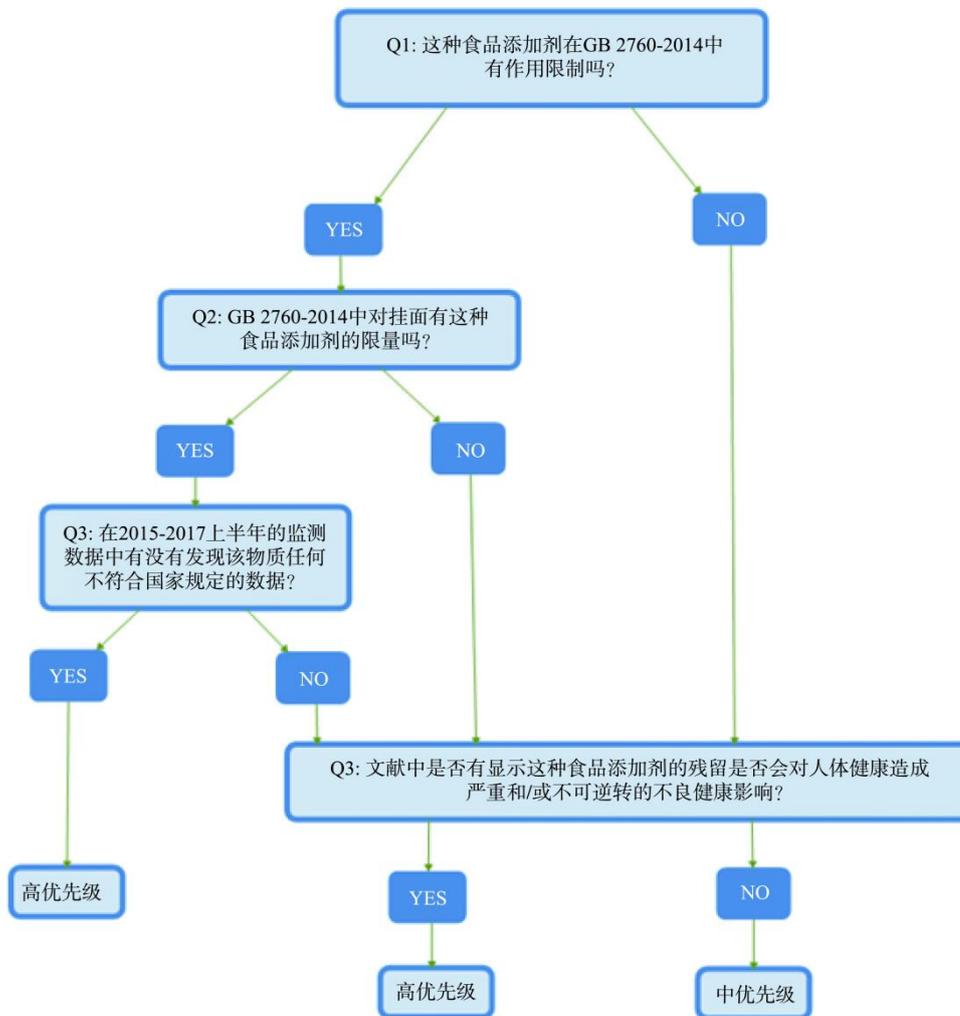


图 2 食品添加剂决策树示意图

Fig.2 Food additive decision tree

本数据中检测项目的添加剂包括二氧化钛、乙二胺四乙酸二钠、柠檬黄、苋菜红、胭脂红、日落黄、诱惑红、亮蓝、铝的残留物。铝、二氧化钛和乙二胺四乙酸二钠在 GB 2760-2014 中有使用限制, 但没有对挂面食品设置使用限量。但文献中表示这种食品添加剂的过量使用会对人体健康造成严重影响, 因此为高优先级。

柠檬黄、日落黄、诱惑红和亮蓝在 GB 2760-2014 中有使用限制, 且对挂面食品设置使用限量。但在 2015~2017 年监测数据中没有发现该物质不符合国标规定的的数据, 文献表示这种食品添加剂的一般使用情况下不会对人体健康造成严重影响, 因此为中优先级。苋菜红、胭脂红在 GB 2760-2014 中有使用限制, 但没有对挂面食品设置使用限量。文献表示这种食品添加剂的一般使用情况下不会对人体健康造成严重影响, 因此为中优先级。

3.3 真菌毒素

GB 2761-2017 中规定无论是否制定真菌毒素限量, 食品生产和加工者均应采取控制措施, 使食品中真菌毒素的含量达到最低水平。本数据中检测项目的真菌毒素包括黄曲霉毒素 B₁ 和脱氧雪腐镰刀菌烯醇。对真菌毒素进行研究时, 从数据中提取初 3 种属性, 分别是:

(1)2015~2017 上半年的数据中, 是否存在该物质超标的情况(标为属性 1)

(2)真菌毒素是否常出现在挂面生产原料中(标为属性 2)

(3)真菌毒素的毒性研究是否完善(标为属性 3)

这 3 种属性都是离散属性, 根据现有数据计算出根节点的信息熵为 0.765, 不同属性的信息增益结果见表 3。

表 3 真菌毒素决策树信息增益计算结果
Table 3 Calculation results of mycotoxins decision tree information gain

属性	信息增益
1	0.202
2	0.101
3	0.194
4	0.173

属性 1 的信息增益最大, 因此选定属性 1 为根节点。依次计算信息增益, 构建叶子节点, 最终的决策树如图 3。

本数据中检测项目的真菌毒素包括黄曲霉毒素 B₁ 和脱氧雪腐镰刀菌烯醇。黄曲霉毒素 B₁ 和脱氧雪腐镰刀菌烯醇在 2015~2017 年上半年的数据中未出现黄曲霉毒素 B₁ 过量导致挂面产品不合格的情况, 但这两种真菌毒素易出现在生产原料如小麦粉中, 因此分类为最高级。

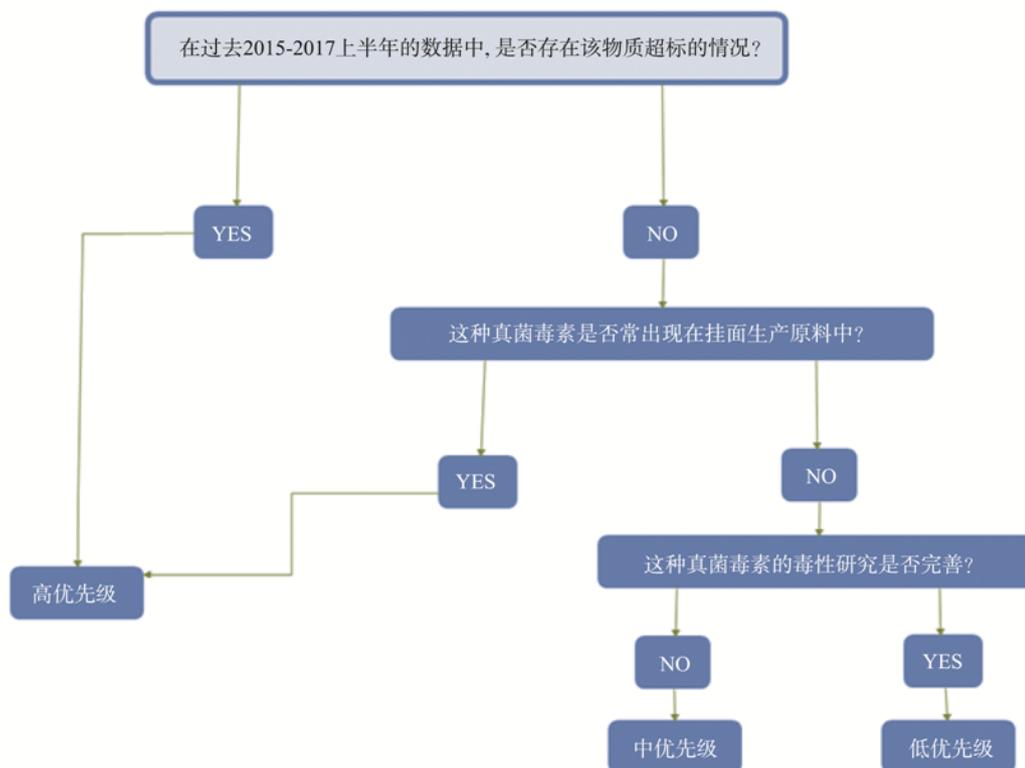


图 3 真菌毒素决策树示意图
Fig.3 Mycotoxins decision tree

4 结论与讨论

本研究基于挂面的检测数据,将检测项目分为污染物、食品添加剂和真菌毒素 3 类,通过计算分别获得信息增益值进而确定根节点和叶子节点,最终得到铅、铝、二氧化钛和乙二胺四乙酸二钠、黄曲霉毒素 B₁ 和脱氧雪腐镰刀菌烯醇为高优先级检测物质,而柠檬黄、日落黄、诱惑红和亮蓝、苋菜红、胭脂红为中优先级检测物质。挂面决策树的建立可以用于快速筛查挂面食品安全危害,为监督监测重点对象提供依据,降低检测成本、提高检测效率。该方法具有可拓展性,可以利用最新监测数据,对其他类型粮食的风险等级进行评估,得到其潜在危害物质的优先次序物质清单。

参考文献

- [1] 叶舟舟, 许晓岚, 生吉萍. 公众对食品安全风险预警信息的信任水平与影响因素研究[J]. 农产品质量与安全, 2018, 94(4): 43-48.
Ye ZZ, Xu XL, Sheng JP. Research on public trust level and influencing factors of food safety risk warning information [J]. Agric Prod Qual Saf, 2018, 94(4): 43-48.
- [2] 陈君石. 食品安全风险评估概述[J]. 中国食品卫生杂志, 2011, 23(1): 4-7.
Chen JS. Overview of food safety risk assessment [J]. Chin J Food Hyg, 2011, 23(1): 4-7.
- [3] 陈君石. 风险评估在食品安全监管中的作用[J]. 农产品质量与安全, 2009, (3): 4-8.
Chen JS. The role of risk assessment in food safety regulation [J]. Agric Prod Qual Saf, 2009, (3): 4-8.
- [4] 唐晓纯. 国家食品安全风险监测评估与预警体系建设及其问题思考[J]. 食品科学, 2013, 34(15): 342-348.
Tang XC. Thoughts on the construction of national food safety risk monitoring, evaluation and early warning system [J]. Food Sci, 2013, 34(15): 342-348.
- [5] 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.
Zhou SJ, Dun ZJ, Liang JH, *et al.* Research on food risk grading method based on semi-quantitative risk assessment [J]. Chin J Food Hyg, 2015, 27(5): 576-585.
- [6] 朱江辉, 宋筱瑜, 王晔茹, 等. 食品微生物风险分级研究进展[J]. 中国食品卫生杂志, 2015, 27(3): 322-329.
Zhu JH, Song XY, Wang YR, *et al.* Progress in research on food microbial risk grading [J]. Chin J Food Hyg, 2015, 27(3): 322-329.
- [7] 谢姐姐. 决策树算法综述[J]. 软件导刊, 2015, 14(11): 63-65.
Xie NN. Summary of decision tree algorithm [J]. Software Guide, 2015, 14(11): 63-65.
- [8] 杨娜. 云南省粮食安全预警系统的设计与实现[D]. 昆明: 云南大学, 2010.
Yang N. Design and implementation of Yunnan province food security early warning system [D]. Kunming: Yunnan University, 2010.
- [9] 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.
Ji SQ, Wu Fang, Li NX. Static early warning model of vegetable diseases based on decision tree [J]. J Tianjin Agric Coll, 2017, 24(2): 77-80.
- [10] 陆荣展, 相秉仁, 徐建平. 决策树算法在药品 GSP 预警管理中的应用[J]. 医学信息学杂志, 2009, 30(5): 29-31.
Lu RZ, Xiang BR, Xu JP. Application of decision tree algorithm in GSP early warning management [J]. J Med Inform, 2009, 30(5): 29-31.
- [11] Van DFHJ, Van AED, Raley M, *et al.* Critical review of methodology and application of risk ranking for prioritisation of food and feed related issues, on the basis of the size of anticipated health impact [J]. EFSA Support Publ, 2015, 12(1): 1-106.
- [12] GB 2760-2014 食品安全国家标准 食品添加剂使用标准[S].
GB 2760-2014 National food safety standard-Use of food additive [S].
- [13] GB 2761-2017 食品安全国家标准 食品中真菌毒素限量[S].
GB 2761-2017 National food safety standard-Limit of mycotoxins in foods [S].
- [14] GB 2762-2017 食品安全国家标准食品中污染物限量[S].
GB 2762-2017 National food safety standard-Limit of pollutant in foods [S].

(责任编辑: 李磅礴)

作者简介



生吉萍, 教授, 博士生导师, 主要研究方向为食品安全管理与食品经济管理。
E-mail: shengjiping@126.com