

# 基于大数据挖掘的医院食品安全风险预测 管理模型

杨丽娟<sup>1\*</sup>, 林 健<sup>2</sup>

(1. 无锡市太湖医院, 无锡 214000; 2. 无锡商业职业技术学院旅游管理学院, 无锡 214000)

**摘要:** **目的** 设计基于大数据挖掘的医院食品安全风险预测管理模型。**方法** 建立医院食品安全风险评估指标体系, 并采用三标度法比较影响食品安全指标的重要程度, 得到各个指标的权重, 后计算食品样品中检测指标不合格率与不合格数, 将检测指标的不合格率与不合格数结合为一个风险系数, 完成对医院食品安全风险预测。**结果** 实验证明, 此次设计的基于大数据挖掘的医院食品安全风险预测管理模型比传统模型预测准确度高。**结论** 可为食品安全领域的发展提供参考依据。

**关键词:** 大数据挖掘; 食品安全; 风险预测

## Risk prediction management model of hospital food safety based on big data mining

YANG Li-Juan<sup>1\*</sup>, LIN Jian<sup>2</sup>

(1. Wuxi Taihu Hospital, Wuxi 214000, China; 2. Institute of Tourism Management, Wuxi Commercial Vocational and Technical College, Wuxi 214000, China)

**ABSTRACT: Objective** To design a risk prediction management model of hospital food safety based on big data mining. **Methods** The index system of hospital food safety risk assessment was established. The three-scale method was used to compare the importance of influencing food safety indexes, and the weight of each index was obtained. After that, the unqualified rate and unqualified number of detection indexes in food samples were calculated, and the unqualified rate and unqualified number of detection indexes were combined into a risk coefficient to complete the risk prediction of hospital food safety. **Results** The experiment proves that the hospital food safety risk prediction and management model designed based on big data mining is more accurate than the traditional model. **Conclusion** It can provide reference for the development of food safety field.

**KEY WORDS:** big data mining; food safety; risk prediction

## 1 引言

医院食品安全风险种类较多、纷繁复杂, 如何对其有效预测受到了社会各界的高度重视。精准预测医院食品安全风险是进行食品安全风险评估与管理的前提, 但医院食

品安全风险观测较难。传统的医院食品安全风险预测管理方法虽然实现较为容易, 也能够得到一定的预测结果, 但该方法不能很好的对医院食品安全风险因子的变化趋势分析, 风险预测准确度较低, 为此设计一种基于大数据挖掘的医院食品安全风险预测管理模型。大数据挖掘是将潜在的、隐含的信息从庞大的数据中挖掘出来, 从大量、不完

\*通讯作者: 杨丽娟, 中级经济师, 主要研究方向为人力资源管理、医院管理。E-mail: wangzyuan1988@163.com

\*Corresponding author: YANG Li-Juan, Intermediate Economist, Wuxi Taihu Hospital, Wuxi 214000, China. E-mail: wangzyuan1988@163.com

程。能够从大量数据中挖掘以及发现数据中的规律与趋势,为此,将大数据挖掘技术应用到医院食品安全风险预测管理中具有重要意义。本研究设计了基于大数据挖掘的医院食品安全风险预测管理模型,以期为食品安全管理提供支撑条件。

## 2 医院食品安全风险评估指标体系构建

结合国家食品安全标准以及生产规范标准,采用大数据挖掘技术挖掘医院食品数据,分析食品安全供应链中原料生产供应环节<sup>[1]</sup>、加工环节、运输环节以及消费环节中涉及的各项指标。建立的医院食品安全风险评估指标体系<sup>[2]</sup>,如表 1 所示:

分析上表可知,医院食品安全风险评估体系种类多样,因此通过大数据挖掘技术挖掘医院食品各个要素<sup>[3]</sup>之间的内在关系,建立食品安全指标权重,采用三标度法<sup>[4]</sup>比较影响食品安全指标的重要程度,如下所示:

$$a_{ij} = \begin{cases} 0, & i \text{ 元素 } j \text{ 元素重要} \\ 1, & i \text{ 元素 } j \text{ 元素同等重要} \\ 2, & i \text{ 元素比 } j \text{ 元素重要} \end{cases} \quad (1)$$

得到其中各层指标因素的比较矩阵  $A$ , 如下图所示:

$$A=(a_{ij}) = \begin{bmatrix} a_{11}, a_{12}, \dots, a_{1n} \\ a_{21}, a_{22}, \dots, a_{2n} \\ \vdots \\ a_{n1}, a_{n2}, \dots, a_{nm} \end{bmatrix} \quad (2)$$

依据上述矩阵,计算指标中各个因素<sup>[5]</sup>的排序指数,计算公式如下:

$$R = \sum_{i=1}^n a * h \quad (3)$$

公式(3)中,  $R$  代表指标中的最大特征值,  $\sum_{i=1}^n a$  代表最大特征值的特征向量,  $h$  代表指标数量。

依据上述过程,完成医院食品安全风险评估指标体系的构建,并对每个指标建立起相应的权重,为医院食品安全风险预测提供基础依据。

## 3 医院食品安全风险系数计算

依据上述评估指标,计算食品样品中测试参数的不合格率<sup>[6]</sup>,计算公式为:

$$x_u = \frac{\sum d / u}{n_t} \quad (4)$$

公式(4)中,  $x_u$  代表食品中的第  $u$  类待测参数,  $n_t$  为待测参数总数,  $\sum d$  为待测参数的食品批次总数。

表 1 医院食品安全风险评估指标体系  
Table 1 Hospital food safety risk assessment index system

一级	二级	三级
原料生产供应	种植	农药、化肥
		土壤
		水质
		温湿度
		微生物
食品加工生产	运输	检验方法
		加工设备
		卫生条件
		添加剂
		湿度、温度
运输	存储	材料
		微生物残留
		湿度、温度
		湿度、温度
		湿度、温度
消费	运输条件	微生物
	存储条件	湿度、温度
	零售环境	保质期

将所有检测的食品批次集合<sup>[7]</sup>中, 得到食品中第  $i$  类待测参数的不合格率, 表达式为:

$$x_i = \frac{h}{\sum c/n_i} \quad (5)$$

式中,  $x_i$  表示第  $i$  类待测参数的不合格率,  $\sum c$  为食品批次的不合格率,  $h$  代表不合格的批次总数。

统计医院食品中所有批次集合中每个批次的不合格数, 这类食品的所有集合中, 每个批次都会对应一个不合格数, 将这些食品所有批次集合中的不合格批次总数<sup>[8]</sup>, 计算公式为:

$$m = \sum_{o=1} e/y(x_i + x_u) \quad (6)$$

公式(6)中,  $m$  代表不合格批次总数,  $\sum_{o=1} e$  代表统计参数,  $y$  代表每个产品批次对应的不合格数值。

将检测的食品集合中不合格批次总数除以批次数, 得到该类食品批次总数的不合格率<sup>[9]</sup>, 计算公式为:

$$v = \frac{\sum f}{n(x_u + x_i)} \quad (7)$$

公中,  $v$  代表食品总体不合格率,  $n$  纳入风险评价的食品类别,  $\sum f$  代表每个抽样产品的抽样数量。

将上述计算结果与风险评估指标体系结合, 相关部门就能够预测食品中的危害物<sup>[10]</sup>, 并且能够预测到具体的某一个项目中。

#### 4 医院食品安全风险预测管理模型构建

由于在实际运算时采用 2 个指标计算, 在计算过程中会增加计算时间, 因此将上述提到的不合格率与不合格总数结合, 合成一个单一食品的安全风险系数<sup>[11]</sup>。构建食品安全风险系数的坐标空间<sup>[12]</sup>, 平面坐标上的原点代表检测参数检测值合格的状态, 针对不合格率和不合格度的二维空间<sup>[13]</sup>, 使用坐标点搭配原点距离构建风险系数, 食品风险系数评价结果计算公式如下所示:

$$K_j = \sqrt{x_i + y_i} \quad (8)$$

公式(8)中,  $K_j$  代表待测食品的风险系数<sup>[14]</sup>,  $x_i$  代表数据来源的食品风险系数,  $y_i$  代表综合评价信息。

基于以上计算, 构建的医院食品安全风险预测管理模型如图 1 所示:

医院食品安全风险预测管理模型中, 信息分析层主要汇集检测数据与舆情数据<sup>[15]</sup>, 包括食品企业、行业协会、食品管理部门网站等信息, 将其汇总到模型中; 数据处理层主要为后续风险预测准备基础数据; 数据汇集层, 对预测结果进一步分析, 反应、报道食品预测结果的新动态、新成果与新的消息。

为进一步提高医院食品安全风险预测管理水平, 此次研究的模型中包含医院食品安全风险预测方法以及管理对策。上述完成风险预测方法的设计, 下一步提出管理对策, 如下所示。

加强监管力度, 建立合理的管理机制, 加强自身产品质量安全的管理, 改变单一化的政府管理方式。并考虑食品供应链中相关利益主体的博弈规则, 设计有效的风险控制机制, 完善食品质量安全信息管理手段和技术, 从而连通食品生产、流通与消费领域, 为医院食品安全风险预测上提供基础信息。并构建安全信息体系、信用评价指标以及评价体系, 从而为医院食品安全风险信息化管理提供基础, 以此完成基于大数据挖掘的医院食品安全风险预测管理模型的构建。

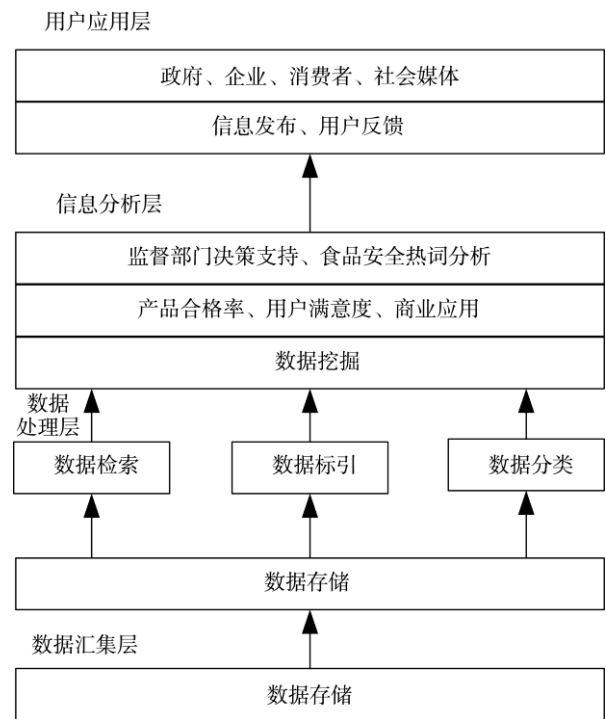


图 1 医院食品安全风险预测管理模型  
Fig.1 Food safety risk prediction management model in hospitals

#### 5 实验对比

为验证上述设计的基于大数据挖掘的医院食品安全风险预测管理模型具备实际应用意义, 进行实验, 并将传统的风险预测管理模型与此次设计模型对比, 对比两种模型的预测准确度。

##### 5.1 实验环境部署

此次实验环境为 3 台网卡的 PC 机、一个百兆带宽网络的集线器与连接网线, 各个节点的配置如表 2 所示:

此次实验搭建的 Hadoop 集群运行与 Linux 环境中,

因此利用虚拟机模拟 Linux 环境, 为保证程序的正常开发, 安装 JDK, 实验中用到的各类软件、版本以及详细说明如表 3 所示:

## 5.2 实验内容来源

实验中采用的实验数据为某相关网站中食品安全事件的相关新闻, 如表 4 所示:

分析上表可知, 这些实验数据内容随意性较大, 文本长度不一, 有效信息差距悬殊, 重复新闻较多, 为此对文本向量化处理, 为实验做准备。上述 8 个事件中, 实际的安全事件发生次数为已知, 分别使用传统模型与此次设计模型对上述数据预测。

## 5.3 实验结果分析

利用传统的基于大数据挖掘的医院食品安全风险预

测管理模型与此次设计的模型对 8 个医院食品安全事件进行用测, 二者预测结果的对比图如图 2 所示:

分析上述对比结果可知, 此次设计的基于大数据挖掘的医院食品安全风险预测管理模型在这 8 个安全事件预测上, 预测结果与实际结果相差较小, 预测准确度高。这是因为此次设计的模型充分建立医院食品安全风险评估指标体系, 并利用大数据挖掘技术, 分析各个指标中潜在的内在关系, 并分析食品中数据不合格率与不合格数, 从而提高了的预测准确度。传统的医院食品安全风险预测管理模型在第 2 个、第 5 个事件预测上, 预测准确度相对较高, 其余几个事件的预测准确度都较低。整体对比结果可知, 传统预测模型的预测准确度都低于此次设计的预测管理模型的准确度。

表 2 3 台 PC 机硬件配置  
Table 2 Hardware configuration of 3 PCS

PC 机	内存大小	硬盘容量
PC1	8.00 GB	932 GB
PC2	1.96 GB	220 GB
PC3	1.75 GB	465 GB

表 3 实验软件环境配置  
Table 3 Experimental software environment configuration

软件名称	软件版本	软件说明
centos	centos6	操作系统软件
hadoop	hadoop-1.1.2	Hadoop 软件
jdk	jdk-6u24	开发编译软件
winscp	winscp 5.1.6	文件管理软件
mabout	mabout-distribution-0.9	机器学习算法库软件

表 4 食品安全事件  
Table 4 Food safety incidents

安全事件发生次数	关键词	聚类描述
54	大米、含量、超标、抽检、调查、中毒	粮食
26	酒、假、生产、销售、白酒、原料、造假	酒
35	蛋白、产品、报告、生物、普通、健康	无法得出
12	食品、作坊、熏、处理、记者、刺鼻、颜色	无法得出
25	毒、豆芽、检测、药、催生、蔬菜	蔬菜
52	油、地沟、作坊、食用油、市场、臭	食用油
36	肉、肯德基、食用油、原料、快餐、肉类	肉类
32	奶粉、含量、婴幼儿、蛋白质、奶制品	奶制品

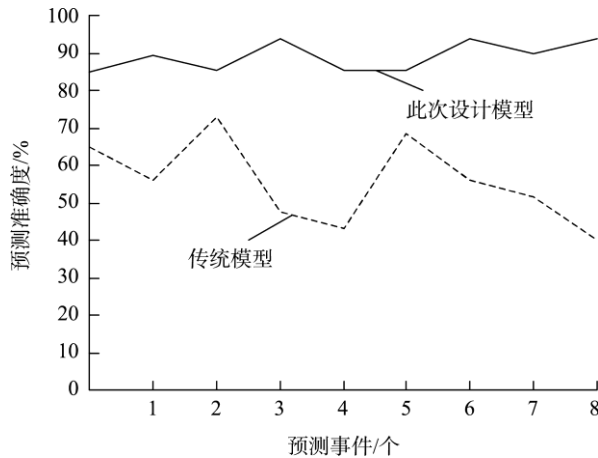


图 2 预测准确度对比

Fig.2 Comparison of prediction accuracy

## 6 结 论

综上所述, 此次研究的基于大数据挖掘的医院食品安全风险预测管理模型在风险识别与预测管理上都取得了丰富的研究成果。为后续食品安全管理的后续研究上提供了丰富的素材与文献支撑, 但是此次研究还存在一定的不足, 这也为进一步研究提供了可以创新和延伸的空间。此次研究的模型中食品数据来源是通过人工收集的方式, 这种数据收集方式样本规模有限, 得到的结果会对预测结果产生一定的偏差。部分结果受到人为因素的影响, 在这一方面有待优化。同时, 各个产品类别的指标体系具有一定的差异性, 只有根据不同产业的特性采用针对性的风险预测方法才能得到更准确的预测结果。同时, 在后续研究中深化研究食品质量安全风险管理机制, 从多个方面对食品安全风险管理, 进一步提高医院食品安全。

## 参考文献

- 孙晓荣, 周子健, 刘翠玲, 等. 光谱预处理结合模拟退火算法的小麦粉面筋含量检测[J]. 食品科学, 2018, 39(2): 222-226.  
Sun XR, Zhou ZJ, Liu CL, *et al.* Near infrared spectroscopic detection of gluten content in wheat flour based on spectral pretreatment and simulated annealing algorithm [J]. Food Sci, 2018, 39(2): 222-226.
- 裴磊, 陈庭强, 王磊, 等. 江苏省食品安全监管信息透明度状况评价研究[J]. 中国调味品, 2018, 43(1): 168-174.  
Pei L, Chen TQ, Wang L, *et al.* Evaluation research on food safety supervision information transparency in Jiangsu province [J]. Chin Condiment, 2018, 43(1): 168-174.
- 王小丹, 杨欣, 徐海滨, 等. 中国不同地区居民谷类食物脱氧雪腐镰刀菌烯醇暴露量及健康风险评估[J]. 中华预防医学杂志, 2019, 53(4): 394-397.  
Wang XD, Yang X, Xu HB, *et al.* Exposure status and health risk assessment of deoxynivalenol from cereals in Chinese population in different regions [J]. Chin J Pre Med, 2019, 53(4): 394-397.
- 牛犇, 洪斌, 穆丽丽, 等. 荧光定量 PCR 在预测微生物学中的应用[J]. 食品工业科技, 2019, 40(5): 275-282.  
Niu B, Hong B, Mu LL, *et al.* Application of fluorescent quantitative pcr in predictive microbiology [J]. Sci Technol Food Ind, 2019, 40(5): 275-282.
- 宋波, 王佛鹏, 周浪, 等. 广西镉地球化学异常区水稻籽粒镉含量预测模型研究[J]. 农业环境科学学报, 2019, 38(12): 2672-2680.  
Song B, Wang FP, Zhou L, *et al.* Prediction model for cadmium concentrations in rice grain under the geochemical background of a cadmium anomaly area in Guangxi [J]. J Agro-Environ Sci, 2019, 38(12): 2672-2680.
- 戚成, 吴雪扬, 陈元美, 等. 肠炎沙门菌的温度、pH 值和水分活度主参数模型的构建[J]. 食品科学, 2018, 39(14): 138-144.  
Qi C, Wu XY, Chen YM, *et al.* Modelling salmonella enteritis growth as a function of temperature, ph, and water activity with cardinal parameter models [J]. Food Sci, 2018, 39(14): 138-144.
- 廖振宇, 曹东丽, 甄鹏, 等. 冷冻饮品质量安全风险分析及防控措施[J]. 食品研究与开发, 2018, 39(13): 210-214.  
Liao ZY, Cao DL, Zhen P, *et al.* The risk analysis of quality safety and the measures of control in frozen drinks [J]. Food Res Dev, 2018, 39(13): 210-214.
- 任美英. 基于整体性治理理论的食品安全监管协同机制研究[J]. 食品与机械, 2019, 35(4): 72-75.  
Ren FY. Research on cooperative mechanism of food safety supervision in China [J]. Food Mach, 2019, 35(4): 72-75.
- 张茂俊, 顾一心, 李颖, 等. 空肠弯曲菌、结肠弯曲菌检验方法团体标准解读[J]. 中华流行病学杂志, 2019, 40(9): 1052-1054.  
Zhang MJ, Gu YX, Li Y, *et al.* Interpretation for the group standards of the Isolation and identification of *Campylobacter jejuni* and *Campylobacter coli* [J]. Chin J Epidemiol, 2019, 40(9): 1052-1054.
- 魏军晓, 耿元波, 岑况. 北京市售大米重金属含量监测及膳食风险评估[J]. 现代食品科技, 2018, 34(6): 267-273.  
Wei JX, Geng YB, Cen K. The heavy metal content monitoring and dietary risk assessment of commercial rice in Beijing [J]. Mod Food Sci Technol, 2018, 34(6): 267-273.
- 谭云龙, 申泽薇, 余灿清, 等. 中国成年人文化程度与体格指标的长期变化关系[J]. 中华流行病学杂志, 2019, 40(1): 26-32.  
Tan YL, Shen ZW, Yu CQ, *et al.* Relationship between educational level and long-term changes of body weight and waist circumference in adults in China [J]. Chin J Epidemiol, 2019, 40(1): 26-32.
- 王伟, 李凤琴. 金黄色葡萄球菌中可移动遗传元件与耐药传播机制研究进展[J]. 中华预防医学杂志, 2018, 52(10): 1067.  
Wang W, Li FQ. Overview of *Staphylococcus aureus* mobile genetic elements and horizontal gene transfer of antimicrobial resistance genes [J]. Chin J Pre Med, 2018, 52(10): 1067.
- 张秀尧, 蔡欣欣, 张晓艺, 等. 二维高效液相色谱-三重四极杆/复合线性离子阱质谱联用法快速测定鸡肉和鸡蛋中利巴韦林总残留量[J]. 质谱学报, 2018, 39(4): 448-456.  
Zhang XY, Cai XX, Zhang XY, *et al.* Rapid determination of total residues of ribavirin in chicken and egg by two-dimensional high performance liquid chromatography-triple quadrupole/linear ion trap mass spectrometry [J]. J Chin Mass Spectrom Soc, 2018, 39(4): 448-456.
- 郭蓉, 王玮, 李敏, 等. 陕西省市售食用植物油及油脂类食品中脂肪酸

氯丙醇酯的污染水平调查与暴露风险评估[J]. 卫生研究, 2019, 48(3): 493-498.

Guo R, Wang W, Li M, *et al.* Contamination and dietary exposure assessment of fatty acid esters of chloropropanediols in vegetable oil and fat-rich food in Shaanxi province [J]. J Hyg Res, 2019, 48(3): 493-498.

[15] 王欣梅, 肖革新, 梁进军, 等. 空间统计学在食品污染物分布研究中的应用[J]. 中华流行病学杂志, 2019, 40(2): 241-246.

Wang XM, Xiao GX, Liang JJ, *et al.* Application of spatial statistics in studying the distribution of food contamination [J]. Chin J Epidemiol, 2019, 40(2): 241-246.

(责任编辑: 王 欣)

### 作者简介



杨丽娟, 中级经济师, 主要研究方向为人力资源管理、医院管理。

E-mail: wangzyuan1988@163.com

## “发酵技术在食品中的应用及安全性评价”专题征稿函

作为众多食品种类的一种, 发酵食品因其独特的风味受到消费者的普遍欢迎。发酵是一种传统的食品储存与加工方法, 是指利用有益微生物加工制造的一类食品, 包括发酵乳制品、酒类、泡菜、酱油、食醋、豆豉等。由于其独特的加工方式, 发酵食品或存在一定的安全隐患, 可能会影响人体健康。

鉴于此, 本刊特别策划“发酵技术在食品中的应用及安全性评价”专题, 特别邀请江南大学生物工程学院吴敬教授担任专题主编, 围绕(1) 菌种的选育和保藏; (2) 发酵工艺的条件优化, 发酵机制, 发酵工程动力学; (3) 发酵食品的分析与检测; (4) 发酵食品的安全性评价及风险评估类等问题展开讨论, 计划在 2020 年 6~7 月出版。

鉴于您在该领域的成就, 学报主编国家食品安全风险评估中心吴永宁研究员及专题主编吴敬教授特别邀请有关食品领域研究人员为本专题撰写稿件, 以期进一步提升该专题的学术质量和影响力。综述及研究论文均可, 请在 2020 年 05 月 25 日前通过网站或 E-mail 投稿。我们将快速处理并经审稿合格后优先发表。

同时烦请您帮忙在同事之间转发一下, 希望您能够推荐该领域的相关专家并提供电话和 E-mail。再次感谢您的关怀与支持!

投稿方式(注明 2020 专题:发酵技术在食品中的应用及安全性评价)

网站: www.chinafoodj.com(备注: 投稿请登录食品安全质量检测学报主页-作者登录-注册投稿-投稿选择“2020 专题:发酵技术在食品中的应用及安全性评价”)

邮箱投稿: E-mail: jfoods@126.com(备注: 2020 专题:发酵技术在食品中的应用及安全性评价专题投稿)

《食品安全质量检测学报》编辑部