

基于机器学习的食品安全风险预警研究现状与展望^{*}

陈夏威 王博远

岑应健 李 笑

(中山市疾病预防控制中心 中山 528403) (广东药科大学 广州 510000)

〔摘要〕 介绍机器学习基本思想、分类与常用方法以及发展历史, 阐述机器学习在食品安全风险预警方面应用现状, 包括 BP 神经网络、贝叶斯网络、自回归求和移动平均模型及决策树方法、支持向量机模型等, 分析当前发展中存在的问题并讨论未来发展方向。

〔关键词〕 食品安全; 食源性疾病; 预警; 机器学习; 健康大数据

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2019.03.011

Current Status and Prospect of the Study on Food Safety Risk Early Warning on the Basis of Machine Learning CHEN Xia-wei, WANG Boyuan, Zhongshan Center for Disease Control and Prevention, Zhongshan 528403, China; CEN Yingjian, LI Xiao, Guangdong Pharmaceutical University, Guangzhou 510000, China

〔Abstract〕 The paper introduces the basic idea, classification, common method and development history of machine learning, elaborates on the application status of machine learning in food safety risk early warning, including BP neural network, Bayesian network, Autoregressive Integrated Moving Average (ARIMA) model and Decision Trees (DT), Support Vector machines (SVM) model, etc. It also analyzes the current issues in the development and discusses the future direction.

〔Keywords〕 food safety; food-borne diseases; early warning; machine learning; health big data

1 引言

食品是人类生存与发展的基础, 然而随着近年

来经济社会的高速发展, 食品安全问题也随之接踵而来, 严重影响人们生活环境, 甚至威胁其生命健康^[1]。媒体的曝光揭露出食品原料质量不合格、加工过程与环境卫生状况差、胡乱添加化学添加剂、食品在运输保存过程中不当操作、政府监管不到位和食品安全法律体系不完善等一系列问题, 这些都是导致食品安全问题不断扩大和食源性疾病频频暴发的原因。当前缺少对于政府部门、医院和相关机构所收集和上报的关于食品安全问题数据信息的全面分析、解读的机器学习方法, 因而无法从数据信息中找出导致食源性疾病暴发的关键特征和各种因子之间的关联, 难以对食品安全问题发展趋势进行正确预测与及时控制。因此有必要引入以 BP 神经

〔修回日期〕 2018-11-13

〔作者简介〕 陈夏威, 副主任医师, 发表论文 9 篇; 通讯作者: 王博远, 硕士, 工程师。

〔基金项目〕 中山市社会公益科技研究专项“基于时空分析模型的中山市食源性疾病和重点食品污染风险预警关键技术研究”(项目编号: 2018B1049); 国家重点研发项目——食品安全社会共治信息技术研究与应用示范“食品安全管控多维动态关联分析技术研究”(项目编号: 2017YFC1602002)。

网络、贝叶斯网络、决策树以及支持向量等为代表的优秀机器学习方法,构建新型综合性食品安全预警模型,在海量而复杂的数据中深度学习找到关键信息和隐藏关系^[2],进而对国家食品安全情况和食源性疾病的流行趋势进行预测预警,最终构建出适应我国国情的食品安全预警体系。

2 机器学习概述

2.1 基本思想

机器学习主要通过计算机模拟人类学习过程,以外环境提供的大量数据信息为对象,对数据信息进行收集、分析、储存、统计,推断出其中规律而建立某种模型,再对未知数据进行推测,无需编译好的程序语言^[3]。简单来说是利用已知部分将未知变为接近实际的已知的过程,属于分析归纳得出的趋势结论。机器学习涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多理论交叉学科,被广泛应用于人工智能、金融与经济数据分析、DNA 生物遗传数据分析等领域^[4],特别在交叉学科领域中起着重要技术支撑的作用^[5]。

2.2 分类和常用方法

2.2.1 分类 机器学习目前大致分为 3 类,分别为监督学习、无监督学习、强化学习^[6]。监督学习由训练资料建立一个程序模型,常用的有人工神经网络、支持向量机、最近邻居法、高斯混合模型、朴素贝叶斯方法、决策树和径向基函数分类等。无监督学习用于在大量无标签数据中进行发掘。常用方式有 K -means 聚类算法和关联规则等。强化学习则可以理解作为一种通过利用试错来找到解决方案的模式,常用有 Q -Learning 以及时间差学习。

2.2.2 常用方法 常用的 10 大机器学习方法有:决策树 (Decision Trees, DT), 朴素贝叶斯分类 (Naive Bayesian Classification, NBC), 最小二乘法 (Ordinary Least Squares Regression, OLSR), 逻辑回归 (Logistic Regression, LR), 支持向量机 (Support Vector Machine, SVM), 集成方法 (Ensemble Methods, EM), 聚类算法 (Clustering Algorithms,

CA), 主成分分析 (Principal Component Analysis, PCA), 奇异值分解 (Singular Value Decomposition, SVD), 独立成分分析 (Independent Component Analysis, ICA)。

2.3 发展历史

2.3.1 概述 1956 年 Arthur Samuel (被称为“机器学习之父”, 1901 - 1990 年) 在 IBM 公司研制可通过大量棋局分析而逐渐识别当前棋局“好棋”与“坏棋”的一个具有自学习、自组织、自适应能力的跳棋程序, 战胜当时的跳棋大师, 模糊的机器学习概念随之产生, 随时间的发展逐渐清晰, 在经历半个多世纪后大致可以将机器学习分为 5 个发展阶段^[7]。

2.3.2 第 1 阶段 20 世纪 40 年代的萌发时期。在这一时期, 心理学家 McCulloch 和数理逻辑学家 Pitts 引入生物学中的神经元概念 (神经网络中的最基本成分), 在分析神经元基本特性的基础上提出 $M-P$ 神经元模型。该模型中每个神经元都能接收到来自其他神经元传递的信号, 这些信号往往经过加权处理再与接受神经元内部的阈值进行比较, 经过神经元激活函数产生对应的输出。

2.3.2 第 2 阶段 20 世纪 50 年代中叶至 60 年代中叶的热烈时期。经典学习规则的提出标志着进入热烈时期, 1957 年美国神经学家 Rosenblatt 提出最简单的前向人工神经网络——感知器, 被广泛应用于文字、声音、信号、记忆识别等领域。

2.3.4 第 3 阶段 20 世纪 60 年代中叶至 70 年代中叶的冷静时期。这一时期由于感知器的单一结构, 局限于处理简单线性可分问题, 加上理论知识的匮乏和计算机有限技术处理, 引来众多学者的质疑, 导致以感知器为核心的人工神经网络的衰败, 机器学习发展停滞不前。

2.3.5 第 4 阶段 20 世纪 70 年代中叶至 80 年代末的复兴时期。1980 年美国卡内基梅隆大学举办首届机器学习国际研讨会, 标志着机器学习在世界范围内的复兴。1986 年加州大学圣地亚哥分校的 Rumelhart 与 McClelland 提出应用于多层神经网络的学习规则即误差逆传播算法 (BP 算法), 引领人工

神经网络发展进入第 2 次高潮, 随后出现自组织映射网络 (Self Organization Mapping, SOM)、竞争型学习网络 (Adaptive Resonance Theory, ART)、径向基函数网络 (Rcecurrent Neural Network, RBF)、级联相关网络 (Cascade Correlation, CC)、递归神经网络 (Radial Basis Function, RNN) 和卷积神经网络 (Convolutional Neural Network, CNN) 等, 机器学习得到发展。

2.3.6 第 5 阶段 20 世纪 90 年代后的多元发展时期。以 SVM 为代表的统计学习迅速发展, 对符号学习的统治地位发起挑战。与此同时, 以通过多个单学习器结合完成学习任务为基本思想的集成学习, 与以通过逐层学习方式解决多隐含层神经网络的初值选择问题, 从而提升分类学习效果, 促使深度学习的提出, 成为机器学习的重要延伸, 使机器学习开始多元化的发展。

3 机器学习在食品安全风险预警中应用现状分析

3.1 BP 神经网络模型

3.1.1 概述 BP (Back Propagation Network) 神经网络是 1986 年由 Rumelhart 和 McClland 为首的科研小组提出的, 是一种按误差逆传播算法训练的多层前馈网络, 其基本思想是梯度下降法, 利用梯度搜索技术使网络的实际输出值和期望输出值的误差均方差为最小, 是目前应用最广泛的神经网络模型之一^[8]。主要的研究应用领域为食品安全预警分析、食品摄入量预测和食品监督抽检合格率预测等。

3.1.2 食品质量预警 章德宾等^[9]利用中国质监系统监督检验部门 2007 年日常检验的原始数据中选取与食品安全最为密切的 167 种检测项目, 按月度划分建立训练样本集作为输入层, 以化学污染、农药残留、兽药残留、重金属、微生物致病菌 5 类作为输出层建立 BP 神经网络食品安全预警模型, 通过初步的应用试验表明 BP 神经网络食品安全预警模型以实际数据训练样本为基础能够进行有效预测, 是一种可行有效的食品安全分析途径。

3.1.3 食品摄入量预测 王霞^[10]以我国 1997 -

2010 年的 9 大类食物摄入量作为训练集进行建模训练, 经过 10 000 次训练使模型误差达到要求, 建立的 BP 神经网络模型对 2011 年食品摄入量进行预测并与实际值对比, 结果表明实验预测值误差非常小且符合应用要求, 再与食品安全评估预警系统结合, 可对食品安全评估的同时还能预测某地某时间段内可能发生的食品安全事故。

3.1.4 食品监督抽检合格率预测 王星云等^[11]利用数据管理平台数据库中食品包括污染物铅的 10 000 条抽检数据作为训练数据集, 选定生产企业省份等 13 个属性作为特征属性, 以合格、不合格作为食品检验结论, 建立 BP 神经网络模型反映食品属性和检验结论之间关系, 运用此关系预测特定属性下检验结论是否合格。这种基于 BP 神经网络的数据挖掘方法能够对食品安全风险进行预警, 为食品安全问题产生提供理论依据。BP 神经网络在拥有高度自学习、自适应能力以及一定的容错能力的同时仍存在一些不足之处, 如需要建立准确率较高的模型, 需要较大的样本量进行训练等, 还需要不断地进行改善优化。

3.2 贝叶斯网络模型

3.2.1 概述 贝叶斯网络 (Bayesian Network, BN) 模型是以动态模型为研究对象的时间序列预测方法, 其不同于一般的统计方法, 不仅利用模型和数据信息, 而且充分利用先验信息, 模式一般为: 先验信息 + 总体分布信息 + 样本信息 → 后验分布信息, 其中还加入决策者的经验和判断信息, 将客观因素和主观因素结合在一起, 对异常情况的发生具有较强的灵活性^[12]。

3.2.2 食源性疾病患者人数预测 在食源性疾病防控过程中, 对于利用当前所掌握数据来确定及预测疫情暴发时间和流行的空间聚集是非常重要的, 这也是学习研究的一个难点。而在得到的数据中存在两种时间延误可能会影响病疫判断, 甚至失去控制疫情关键时间点, 一是患者发病到医生确诊过程中时间的延误, 二是医生确诊到医院上报到相关机构过程中时间的延误。而贝叶斯临近预报模型能够很好地将此类延误影响降低。Wang XueLi 等^[13]根

据国家食品安全风险评估中心食品监测数据库于 2015 年 3 月 1 日 - 2016 年 2 月 28 日湖南省共报告 21 866 例食源性疾病就诊病例中确诊病例 21 226 例, 比例高达 97% 这一数据, 建立贝叶斯层次模型以利用每日来访患者数预测每日真实患者数, 实验通过专业评定表明贝叶斯网络模型具有良好的短期预报性能, 能为今后的食源性疾病监测和控制策略提供方法学依据。

3.2.3 食品供应链污染暴露预测 张丽等^[14]通过对某一食物进行风险分析, 经过裁剪风险因素和分析初始风险事件, 针对食品供应链中物流环节, 以风险因素放射物污染为根风险因素, 建立局部贝叶斯网络模型, 然后进行风险概率评价, 得出放射物污染发生概率是 0.9 时存储仓库污染、运输车污染等概率分别为 0.825、0.785, 发现放射物暴露可能性很大, 表明基于贝叶斯网络模型的食品供应链风险预测可以提前分析发现风险威胁。贝叶斯网络模型需改善的方面包括有减少主观判断或加入专家系统^[15]的建议与评定等, 而继续对贝叶斯网络模型进行改善发掘必将为未来食源性疾病监测和控制策略提供更好的方法与依据, 有助于公共卫生与食品安全的发展。

3.3 ARIMA 模型

3.3.1 概述 食物中毒的发生具有趋势性、周期性、短期波动和明显季度性等特点, 掌握这些特点能大大提高对病情的预测与趋势控制, 而时间序列分析中的自回归求和移动平均模型 (Autoregressive Integrated Moving Average Model, ARIMA) 能为其提供较为准确的参考依据。主要应用为疾病发病率预测。

3.3.2 细菌性痢疾发病率预测 郭建娥等^[16]以太原市疾控中心收集的 1994 - 2003 年痢疾月发病率数据作为训练资料拟合模型, 利用 2004 - 2006 年痢疾发病率数据对模型参数进行修正, 建立预测模型对 2007 - 2008 年太原市痢疾发病率进行预测, 预测结果较好。朋文佳等^[17]以某区 2004 - 2008 年细菌性痢疾月发病率资料作为训练样本建立 ARIMA 模型, 通过 2009 年细菌性痢疾月发病率资料对模型参数进行修正, 预测 2010 年细菌性痢疾月发病率并与实际值进行对比分析, 发现预测值与实际值趋势基本吻合。

3.3.3 食源性疾病发病率预测 万蓉等^[18]以云南省 2004 年 1 月 - 2010 年 12 月食源性疾病月发病率数据作为训练资料建立 ARIMA 模型, 通过 2011 年月发病率资料进行回代, 预测 2012 年食源性疾病月发病率趋势, 结果发现 2011 年各月发病率预测值 95% 可信区间包括真实值, 表明 ARIMA 模型能够对食源性疾病进行短期预测。张爱红等^[19]探讨 ARIMA 模型在食源性疾病发病率预测中的应用, 较好地拟合短期内食源性疾病人数在时间序列上的变化趋势, 是一种短期预测精度较高的预测模型。ARIMA 模型可以较好地拟合食源性疾病的时间变化趋势, 用于预测未来的食源性疾病。但需要注意以下几点: 一是若用于长期预测则需要长期不断地利用监测数据进行模型参数调整, 从而提高模型预测效果; 二是时间序列不宜太短, 否则影响模型的可靠性; 三是若研究地区的发病率发生较大的改变趋势, 需要重新收集积累新的数据, 再对模型进行修正和重新拟合。

3.4 决策树方法

3.4.1 概述 决策树是一种用于决策的工具。每个决策或事件 (即自然状态) 都可能引出两个或多个事件, 导致不同的结果, 将这种决策分支画成图形, 很像一棵树的枝干, 故称决策树。通过树状的逻辑思维方式解决复杂决策问题的一种方法, 即利用信息论原理在相关资料中提取有用信息构建框架, 再以树模型的逻辑性对决策对象进行判断分析。构建决策树的过程一般分为构树、验证剪枝以及基本策树确定。根据专家系统分析训练集, 对训练集中样本进行评价, 利用逻辑关系和算法构建初步决策树, 再利用测试集监测初步决策树, 经过拟合、剪枝与再返测试, 最后优化决策树结构获得成熟决策树。主要的研究应用领域包括农产品安全评估、食品安全情况分析等。

3.4.2 农产品安全评估 赵静娴^[20]针对影响农产品质量安全的数据特点构建基于组合优化决策树的农产品安全评估模型, 通过 2005 - 2009 年北京市、河北省等华北地区的 20 个蔬菜种植基地或区域的调查数据及国家农业科学数据共享中心的数据验证评估模型的有效性, 结果表明基于组合优化决策树

的农产品安全评估模型能够较为准确地对农产品质量安全风险进行评估,具有一定的食品安全风险预警作用。

3.4.3 食品安全情况分析 鄂旭等^[21]对食品安全情况进行分析,对传统决策树进行改进,基于粗糙集变精度模型构建决策树新方法,计算各属性的加权平均变精度粗糙度值,选择值最小的属性作为相应节点,构建具有精确置信度的决策规则。通过食品安全信息表实例证实该方法可行性,且对食品安全决策有更优化的应用。决策树虽然具有将复杂问题分点简化的逻辑思维特性,但产生错误可能性也随之增大,而且无法解决数据噪音问题,所以需要继续改进提高,从而更好地对食品安全进行监管、预警,为食品安全决策提供依据。

3.5 支持向量机模型

3.5.1 概述 SVM 是机器学习的代表方法之一。其与神经网络一样都是从经验数据集中学习的,是数据驱动的学习机。但不同的是神经网络学习的目标函数是经验风险最小化,得到的是小的训练误差,而支持向量机学习的目标函数是结构风险最小化,得到的是小的泛化误差^[22]。在实际应用中支持向量机技术主要用来解决模式识别和函数回归问题,在非线性系统建模、时间序列预测和疾病诊断等领域都有涉及。主要的研究应用领域为粮食安全风险预警、乳制品安全风险评估等。

3.5.2 细菌性食物中毒暴发事件预测预警 段鹤君、邵兵^[23]利用 2001 - 2009 年万方、同方和维普等数据库中的全国范围内正式发表的文献报道中细菌性食物中毒数据资料,快速有效地对细菌性中毒因素进行评估,以数据资料中随机抽取的 70% 数据作为训练集,通过 PCA 和 SVM 建立回归模型,以数据资料中剩余 30% 数据作为测试集进行预测,结果显示达到 80% 以上的预测准确率。

3.5.3 粮食安全风险预警 Zhu Changxing 等^[24]为克服传统食品风险预警模型在假设、样本量、泛化能力等方面存在的不足,将 SVM 分类与预警理论相结合,提出一种基于 SVM 的中国粮食安全风险预警模型。构建 SVM 模型并通过实验计算表明该方法

较传统方法更加精确,可以为决策者提供更可靠的依据。

3.5.4 乳制品安全风险评估 洪雪婷^[25]以乳制品为例,建立多层次、多指标的乳制品安全风险评价指标体系,将实际数据导入基于猫群算法优化的支持向量机模型中进行预测,再通过 Matlab^[26] 比较预测结果,验证模型准确性,表明该模型能够对食品安全进行风险预警,以便更好地对食品质量安全进行监控。SVM 具有结构风险最小化的特点,在解决非线性、小样本、高维的机器学习领域中有独特优势。但是目前 SVM 许多理论解释并不完善,在实际算法中不能完全实现,许多方面仍需要进一步讨论研究。

3.6 当前发展中存在的问题

机器学习的应用需要较多的信息数据,而目前我国食品安全数据的采集、分析和利用仍存在问题^[27]。数据共享机制不完善,上报机制落后。各个监管部门、卫生机构和食品厂商缺乏相互共享的平台,导致信息数据封闭,无法流通利用。此外信息平台落后、情况上报制度不完善,应时所收集的数据没有及时上报有关部门机构,信息无法及时利用导致信息数据滞后^[28]。信息数据质量不高,可利用率低^[29]。利用机器学习建立食品安全风险预警模型需要大量的信息数据作为训练资料,而地方监测哨点发展程度不均匀,导致监测收集的信息数据重复、交叉以及缺乏地区针对性。

4 结语

新时代数据信息快速更新,机器学习已经具备一定解决实际问题的能力,成为一种支持服务类的技术,在各个领域中取得较好成果,但应用于食品安全风险预警中仍需要改进完善。将不同种类的机器学习方法以不同形式组合在一起,跳出以神经网络、贝叶斯以及支持向量机等机器学习为中心的框架,与国内外其他领域中先进的技术相结合,构建新型综合性食品安全预警模型,这必将使机器学习在食品安全预警应用中得到快速发展。大数据时代是社会发展的必然,针对食品安全的大数据特点,如何快速发掘出有

价值的数 据, 掌握其分布规律, 做好预测并防范是未来发展方向。然而面对我国巨大的人口基数、复杂的食品安全环境及消费人群, 需要把握好大数据时代发展契机, 引进新思路、新方法, 加强大数据分析 与利用, 使食品安全与大数据融合, 提高食品安全质量监测强度, 改善人民生活水平^[30]。

参考文献

- 1 旭日干, 庞国芳. 中国食品安全现状、问题及对策战略研究 [M]. 北京: 科学出版社, 2015.
- 2 王雅洁, 杨冰, 罗艳, 等. 大数据挖掘在食品安全风险预警领域的应用 [J]. 安徽农业科学, 2015, 43 (8): 332 - 334.
- 3 Michie D, Spiegelhalter D J, Taylor C C, et al. Machine Learning, Neural and Statistical Classification [J]. Technometrics, 1994, 37 (4): 459.
- 4 张平. 安全领域中机器学习的对抗和博弈 [J]. 计算机与网络, 2016, 42 (22): 60 - 62.
- 5 陈康, 向勇, 喻超. 大数据时代机器学习的新趋势 [J]. 电信科学, 2012, 28 (12): 88 - 95.
- 6 张长水, 杨强. 机器学习及其应用 [M]. 北京: 清华大学出版社, 2013.
- 7 周子扬. 机器学习与深度学习的发展及应用 [J]. 电子世界, 2017 (23): 72 - 73.
- 8 Goñi S M, Oddone S, Segura J A, et al. Prediction of Foods Freezing and Thawing Times: artificial neural networks and genetic algorithm approach [J]. Journal of Food Engineering, 2008, 84 (1): 164 - 178.
- 9 章德宾, 徐家鹏, 许建军, 等. 基于监测数据和 BP 神经网络的食品安全预警模型 [J]. 农业工程学报, 2010, 26 (1): 221 - 226.
- 10 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.
- 11 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.
- 12 慕春棣, 叶俊. 用于数据挖掘的贝叶斯网络 [J]. 软件学报, 2000, 11 (5): 660 - 666.
- 13 Wang X, Zhou M, Jia J, 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.
- 14 张丽, 滕飞, 王鹏. 基于贝叶斯网络的食品供应链风险评价研究 [J]. 食品研究与开发, 2014, 35 (18): 179 - 182.
- 15 冯定. 神经网络专家系统 [M]. 北京: 科学出版社, 2006.
- 16 郭建娥, 刘桂芬. 构建山西省太原市痢疾发病率的 ARIMA 模型 [J]. 疾病监测, 2007, 22 (4): 222 - 225.
- 17 朋文佳, 朱玉, 何倩, 等. ARIMA 乘积季节模型在细菌性痢疾月发病率预测中的应用 [J]. 中国卫生统计, 2011, 28 (6): 645 - 647.
- 18 万蓉, 李娟娟, 王晓雯. ARIMA 乘积季节模型在食源性疾 病月发病率预测中的应用 [J]. 昆明医科大学学报, 2012, 33 (6): 48 - 52.
- 19 张爱红, 周培, 申铜倩, 等. 乘积季节 ARIMA 模型在食 源性疾病预测中的应用 [J]. 中国卫生统计, 2014, 31 (1): 68 - 69, 73.
- 20 赵静娴. 基于决策树 的食品安全评估研究 [J]. 安徽农业科学, 2011, 39 (32): 20259, 20302.
- 21 鄂旭, 任俊原, 毕嘉娜, 等. 基于粗糙变精度的食品安全决策树研究 [J]. 计算机技术与发展, 2014, 24 (1): 242 - 245.
- 22 Xuegong Z. Introduction to Statistical Learning Theory and Support Vector Machines [J]. Acta Automatica Sinica, 2000, 26 (1): 32 - 42.
- 23 段鹤君, 邵兵. 基于主成分分析和支持向量机的细菌性食物中毒预警研究 [J]. 卫生研究, 2010, 39 (6): 747 - 750.
- 24 Changxing Z, Fenge W. Study on Risk Pre - warning Model of China Food Based on SVM Classification [C]. Henan: E - product E - service and E - entertainment (ICEEE), 2010 International Conference, 2010.
- 25 洪雪婷. 基于核方法 的食品安全风险评估与预警方法研究 [D]. 天津: 天津科技大学, 2015.
- 26 葛哲学, 孙志强. 神经网络理论与 MATLAB R2007 实现 [M]. 北京: 电子工业出版社, 2007.
- 27 王晓丽. 机器学习实际应用面临的问题及解决思路 [J]. 新教育时代电子杂志 (教师版), 2017 (39): 289 - 290.
- 28 杨会志. 数据挖掘技术的主要方法及其发展方向 [J]. 河北科技大学学报, 2000, 21 (3): 77 - 80.
- 29 肖革新, 肖辉. 基于空间统计的食品安全研究现状与展望 [J]. 中国食品卫生杂志, 2016, 28 (4): 409 - 414.
- 30 肖革新, 肖辉, 刘杨. 食品安全大数据分析思考 [J]. 中国数字医学, 2014, 9 (1): 4 - 7.