

# 食品安全大数据可视化关联分析

陈谊, 孙梦, 武彩霞, 孙小然

北京工商大学计算机学院食品安全大数据技术北京市重点实验室, 北京 100048

## 摘要

随着检测技术的提高和互联网技术的广泛应用,食品安全数据的规模不断增大、类型不断增多,对数据分析技术提出了极大挑战。近年来出现的可视分析技术,通过提供图形交互界面,帮助领域人员深入理解数据并洞悉数据中的隐含规律,提高对食品安全风险的分析、发现、预警和溯源能力,为食品安全监测和管控提供了新手段。首先分析了食品安全数据的主要来源、特征和分析任务;然后提出了一种关联可视分析技术分类方法,从属性关联、实体关联、对比分析和时空分析4个方面阐述了近10年来的食品安全大数据可视化关联分析方法;最后提出了该领域存在的问题和挑战。

## 关键词

可视分析;大数据;食品安全;关联关系

中图分类号:TP391.41

文献标识码:A

doi: 10.11959/j.issn.2096-0271.2021014

## *Visual associations analysis of big data in food safety*

CHEN Yi, SUN Meng, WU Caixia, SUN Xiaoran

Beijing Key Laboratory of Big Data Technology for Food Safety, School of Computer Science and Engineering, Beijing Technology and Business University, Beijing 100048, China

## *Abstract*

With the improvement of detection technology and the wide application of Internet technology, the scale and types of food safety data continue to increase, which poses great challenges to data analysis technology. Visual analysis, which has emerged in recent years, can help domain experts gain a deeper understanding of the data and insight into the hidden patterns in the data by providing a graphical interactive interface. This in turn can improve the detection, analysis, early warning and traceability of food safety risks, providing new tools for food safety monitoring and surveillance. Firstly, the main sources, characteristics and analysis tasks of food safety big data were analyzed. Then, a classification method for visual associations analysis techniques was proposed, and the visual associations analysis methods for food safety big data in the past 10 years were described from four aspects: attribute correlation, entity associations, comparative analysis and spatio-temporal analysis. Finally, the problems and challenges in this field were presented.

## *Key words*

visual analysis, big data, food safety, associated relation

## 1 引言

食品安全关系到人们的身体健康和生命安全,受到世界各国的广泛关注。食品安全问题主要包括病原微生物污染、农药兽药残留、重金属和真菌毒素污染、食品添加剂的非法和掺杂使用<sup>[1]</sup>,是一个非常复杂的问题,涉及从种植养殖、生产加工、运输贮藏到餐桌消费的全过程,为此各国政府部门均加强了对从农田到餐桌全链条的食品安全监测和管控,进而产生了大量的食品安全数据。这些数据涉及食品的种类、营养、污染物、时间、地域等多维度信息,具有多维、时空、层次、关联等特征<sup>[2]</sup>,对这些数据的关联分析是食品安全领域的重要分析任务,包括属性关联、实体关联、对比分析和时空分析。随着检测技术的进步和物联网技术的应用,食品安全数据的数量、类型不断增多,产生的速度不断加快,开始呈现出大数据的特征<sup>[3]</sup>。这一方面为基于数据驱动的食品安全风险分析提供了丰富的数据资源,另一方面也对数据分析技术提出了极大的挑战。

## 2 食品安全数据的来源、特征与关联分析任务

### 2.1 数据来源

食品安全数据来源广泛,主要包括如下几个方面。

- 各类食品安全检测仪器的检测结果。这些检测仪器可以是:用于检测农药残留的色谱-质谱仪、用于监测食品质量的射频识别(radio frequency identification, RFID)传感器和视频设

备、用于快速检测食品安全的移动设备(如手机)等。

- 与食品安全相关的标准文件。如食品中各种危害物(农药残留、重金属、致病菌等)的限量标准、检测方法标准、食品中营养成分的限量标准等。

- 互联网数据。如新闻、微博、Twitter等社交媒体上的相关评论等。

- 在线数据库<sup>[4]</sup>。各国食品安全管理部门或组织都会在线发布数据,这些数据包含与食品安全相关的信息,如各国的相关标准、食品中污染物的抽检结果和分析报告、出入境检验检疫不合格食品信息、食品消费数据、风险预警信息等。

**表1**给出了几个国内外典型的食品安全在线数据库。GEMS/Food(global environment monitoring system-food contamination monitoring and assessment programme)是全球环境监测/食品污染监测和评估系统,它由世界卫生组织(World Health Organization, WHO)发布,包含世界多个区域的食物污染物监测数据;RASFF(rapid alert system for food and feed)是欧盟食品和饲料类快速预警系统,是现在由欧洲联盟委员会(European Commission, EC)公开的经常使用的食品安全在线数据库,该系统可以按照通知、通知类别、危害物、日期、产品、关键字等进行数据筛选。EFSA-Data是由欧洲食品安全局(European Food Safety Authority, EFSA)发布的欧盟地区食品安全数据收集和分析结果,其中包含食品消费、食品成分、生物危害、化学危害、化学污染物、化学残留物、植物学纲要和标准化数据等相关数据,风险评估者可使用该数据集计算出消费者对某种危害的暴露程度,监控食品安全计划的有效性。美国食品药品监督管理局(Food and Drug Administration,

表1 食品安全在线数据库

数据库	数据性质	数据描述	国家/地区	组织
GEMS/Food	全球环境监测/食品污染监测和评估系统	全球各地区食品中污染物(隔、铅、汞、毒素等)的监测结果	全球	世界卫生组织
RASFF	欧盟食品和饲料类快速预警系统	针对欧盟各成员国内部,因食品不符合安全要求或标识不准确等引起的风险报告和潜在问题通报	欧盟	欧洲联盟委员会
EFSA-Data	数据收集、标准化与分析	食品消费、食品成分、生物危害、化学危害、化学污染物、化学残留物等数据	欧盟	欧盟食品安全局
pesticide residue monitoring program reports and data	农药残留监测计划报告和数据库	美国各州食品药品监督管理局农药残留监测计划的结果总结以及按商品类别分类的食品监测数据	美国	美国食品药品监督管理局
FDA recent recalls, market withdrawals & safety alerts	FDA最新召回、市场撤销和安全警报	FDA监管产品召回的信息,包括日期、品牌名称、产品描述、产品类别、召回原因说明、公司等	美国	美国食品药品监督管理局
食品安全抽检结果查询系统	食品安全抽检公布结果查询系统	定期发布国内产品抽检结果,包括食品名称、抽检次数、检测项目、检出污染物及含量、是否合格等信息	中国	国家市场监督管理总局
食品安全国家标准数据检索平台	食品安全国家标准数据检索平台	中国食品安全国家标准公告,各类食品中的危害物及具体限量标准	中国	国家市场监督管理总局
中国海关未准入境食品信息	未准入境食品信息列表	中国海关每个月发布抽检不合格、未准入境的食品信息,包括产品名称、产地、生产企业、未准入境原因、进境口岸等	中国	中国海关总署
各国食品安全标准	各国食品安全相关标准	各国食品中营养成分、污染物等的限量标准	中国	中国海关总署

FDA)发布了一些数据库,如农药残留监测计划报告和数据库、FDA监管产品召回的信息等。中国食品安全国家标准和食品安全抽检结果由国家市场监督管理总局(State Administration for Market Regulation, SAMR)发布,包含我国各污染物的限量标准、抽检结果全部合格和不合格产品信息食品名称和抽检次数等。对于进出口食品安全数据,如各国有关标准和未准入境食品信息由中国海关总署(General Administration of Customs of the People's Republic of China, GACC)发布。此外,中国还有一些食品安全数据集是由商业组织提供的,例如食品伙伴网的专业食品安全数据库,包含安全性指标、食品抽检信息、化学污染、微生物、进出口信息、认证信息、营养数据等。上述数据库大多为用户提供了根据地区、食品分类、危害

物、时间等关键字段进行查询的功能。

## 2.2 数据特征

从数据来源可以看出,食品安全数据开始呈现规模巨大(volume)、类型多样(variety)、产生速度快(velocity)、价值密度低(value)、不确定性(veracity)和动态可变性(variability)的大数据6V特征<sup>[5]</sup>。从数据分析的角度来看,它还具有多源、多维、层次、时空和动态可变等特征。多源是指食品安全数据来源广泛,如前文所述,其可以是检测结果、监测数据、标准文件、监管数据、互联网数据、在线数据库等;多维是指食品安全数据具有多维属性,例如,食品中含有多种营养成分,包含蛋白质、维生素、糖、脂肪等1500多种;层次是指数据具有树形的层次结构,例

如,农产品分类、农药分类、地域的行政区划等都具有层次特征;时空是指食品安全数据有一定的时间属性和空间属性,通常人们需要统计食品安全数据按空间分布和时间分布的态势;动态可变性是指食品安全数据是动态变化的,如监测数据、视频数据、社交媒体数据等。

### 2.3 关联分析任务

食品安全监管的目的是通过相关技术手段进行风险识别、风险分析和风险评估<sup>[6]</sup>,进而实现风险预警和追根溯源,以最大限度地预防、减轻和消除食品安全的危害和风险,确保人民群众的食品安全。随着食品安全大数据的出现,数据驱动的风险分析已成为食品安全决策和监管的重要手段之一<sup>[7]</sup>。通过对食品安全数据的关联分析,领域人员可以掌握数据的分布特征、发现异常、探索数据间的隐含关联,以支持食品安全风险识别、风险评估、风险预警和追根溯源。

根据食品安全监管的需求和数据的特征,笔者将食品安全数据的关联分析分为以下4类:属性关联、实体关联、对比分析和时空分析。

- 属性关联是指属性间的相关性分析。食品安全数据通常是多维或高维数据,涉及食品的各种属性(名称、分类、营养成分等)、危害物的各种属性(名称、类别、毒性、成分等)、多国限量标准等,通过属性间的相关性分析,可以进行营养食品推荐或食品安全风险预测。

- 实体关联是指实体间的关联分析。食品安全数据集中通常含有多个食品或危害物等实体,这些实体间的关联可以是树状的层次关系,也可以是网状的关联关系,通过对这些关联关系的探索,可以帮助发现异常实体,如高风险的食品或危害物。

- 对比分析。作为关联分析的一种特殊形式,对比分析在食品安全领域中普遍存在,如检测结果与限量标准的对比和食品污染程度的排名可以帮助发现不合格食品 and 高风险食品;多国限量标准的对比可以帮助监管部门发现两个国家同类标准的差异,如农药残留最大限量标准,进而改进和完善我国现有标准。

- 时空分析。食品安全数据的许多属性与时间或空间关联,这类关联分析可以使用户了解数据的空间分布和时变特征,从而对食品安全事件进行预测和溯源。

针对上述关联分析需求,当前基于统计学和数据挖掘的方法在一定程度上有效地解决了食品安全数据的关联分析问题,然而这些方法大多是全自动的过程,没有考虑人的经验和知识。随着食品安全大数据时代的到来,食品安全监测预警工作的思维方式和范式将发生根本性的变化,食品安全数据监测预警的分析对象和研究内容更加细化,数据获取技术更加便捷,这就要求信息处理技术更加智能,信息表达和服务技术更加精准<sup>[8]</sup>。

## 3 食品安全数据的关联可视分析方法

近年来出现的可视分析技术将人的经验智慧与机器的运算能力紧密地结合在一起<sup>[9]</sup>,通过提供有效的交互可视界面,帮助人们快速准确地观察、过滤、探索、理解和分析大规模数据,从而有效地发现隐藏在数据内部的特征和规律。数据可视分析流程如下<sup>[10]</sup>:首先对原始数据进行收集和存储,并将其存储到数据文件或数据库中,然后通过数据分析得到分析结果,再将分析结果通过可视化映射形成可视化视图,并呈现给用户。用户则根据其分析任

务,在上述流程中的各个阶段与数据进行交互(如选择和过滤要分析的数据、调节分析模型中的参数、切换不同的视图),以实现<sup>对数据的全方位分析</sup>。这种人在回路的可视分析方法为食品安全数据的关联分析提供了新思路,本节将根据食品安全领域中的4类关联分析需求,分类介绍关联可视分析方法,见表2。

### 3.1 属性间的关联可视分析

食品安全数据属性间的关联分析可以分为两属性相关性分析和多属性相关性分析,通常使用散点图、散点图矩阵、平行坐标、邻接矩阵等可视化方法将数据呈现给用户,通过交互手段,用户可以探索数据属性间的相关性和关联规律。

#### 3.1.1 两属性之间的相关性可视分析

散点图由直角坐标系上的数据点构成,表明数据的二维属性之间的关系。Bian R Z等人<sup>[11]</sup>提出了一种基于隐函数微分的多维投影方法,将红酒数据集通过主成分分析(principal component analysis, PCA)方法投影到二维平面上,

并形成散点图,如图1(a)所示。对于多属性的相关性分析,可以使用散点图的扩展方法——散点图矩阵<sup>[30-31]</sup>。散点图矩阵将 $n$ 维数据每两维组成一个散点图(数据维度即数据属性),再将它们按照一定的顺序组成 $n \times n$ 的矩阵,即散点图矩阵,以揭示 $n$ 维数据中所有维度两两之间的关系。为了同时探索多维数据属性间的相关性,Yuan X R等人<sup>[12]</sup>提出了一种维度投影矩阵技术,用于交互式地探索和分析高维数据子空间,并将其应用于分析食品营养数据集,每个维度代表某种营养素。首先将美国农业部食品数据集经过多维标度分析(multidimensional scaling, MDS)投影在二维平面中,按照此图的聚类特征将维度分为4个互相独立的组,形成维度投影矩阵,以帮助用户探索和分析食品中多个营养素之间的关联关系,将食品按营养素的关联关系进行合理的聚类,如图1(b)所示。

#### 3.1.2 多属性之间的相关性可视分析

平行坐标(parallel coordinates)用一组平行的轴表示数据的属性,用穿过每一条轴线的一条折线表示一个数据对象,平行的轴是等距离分布的垂直线或水平

表2 食品安全数据关联可视分析方法汇总

关联分类	分析任务描述	可视化方法	食品领域应用实例
属性关联	两属性相关性	散点图 <sup>[11]</sup>	食品营养成分间两两成分的相关性分析
	多属性相关性	散点图矩阵 <sup>[12]</sup> 、平行坐标 <sup>[13-14]</sup> 、邻接矩阵 <sup>[15]</sup>	农药残留多个判定指标上值的差异和相关性分析
实体关联	实体间网络关系	节点-链接法 <sup>[16]</sup> 、弦图 <sup>[17]</sup> 、邻接矩阵 <sup>[18]</sup>	农产品实体与农药实体间的多对多网络关系分析
	实体间层次关系	节点-链接树 <sup>[19]</sup> 、树图 <sup>[20-21]</sup> 、放射环 <sup>[22-23]</sup>	农产品各类别、农药类别、地域行政区划等的层次关系分析
对比分析	数值对比	排序(如单属性和多属性排名) <sup>[24]</sup>	各种食品、各地区等受危害物污染程度的排名
	结构对比	并置(树比较方法之一) <sup>[13,21]</sup>	两国家最大残留限量(maximum residues limit, MRL)标准的对比(被转化为两棵树的对比)
时空分析	时间关联	时间线 <sup>[21,25]</sup> 、主题河流 <sup>[26]</sup>	农产品中农药残留超标率随时间的变化情况
	空间关联	统计地图 <sup>[25,27-29]</sup> 、变形地图 <sup>[25,28-29]</sup>	食品中危害物检出率和超标率在地域上分布情况分析

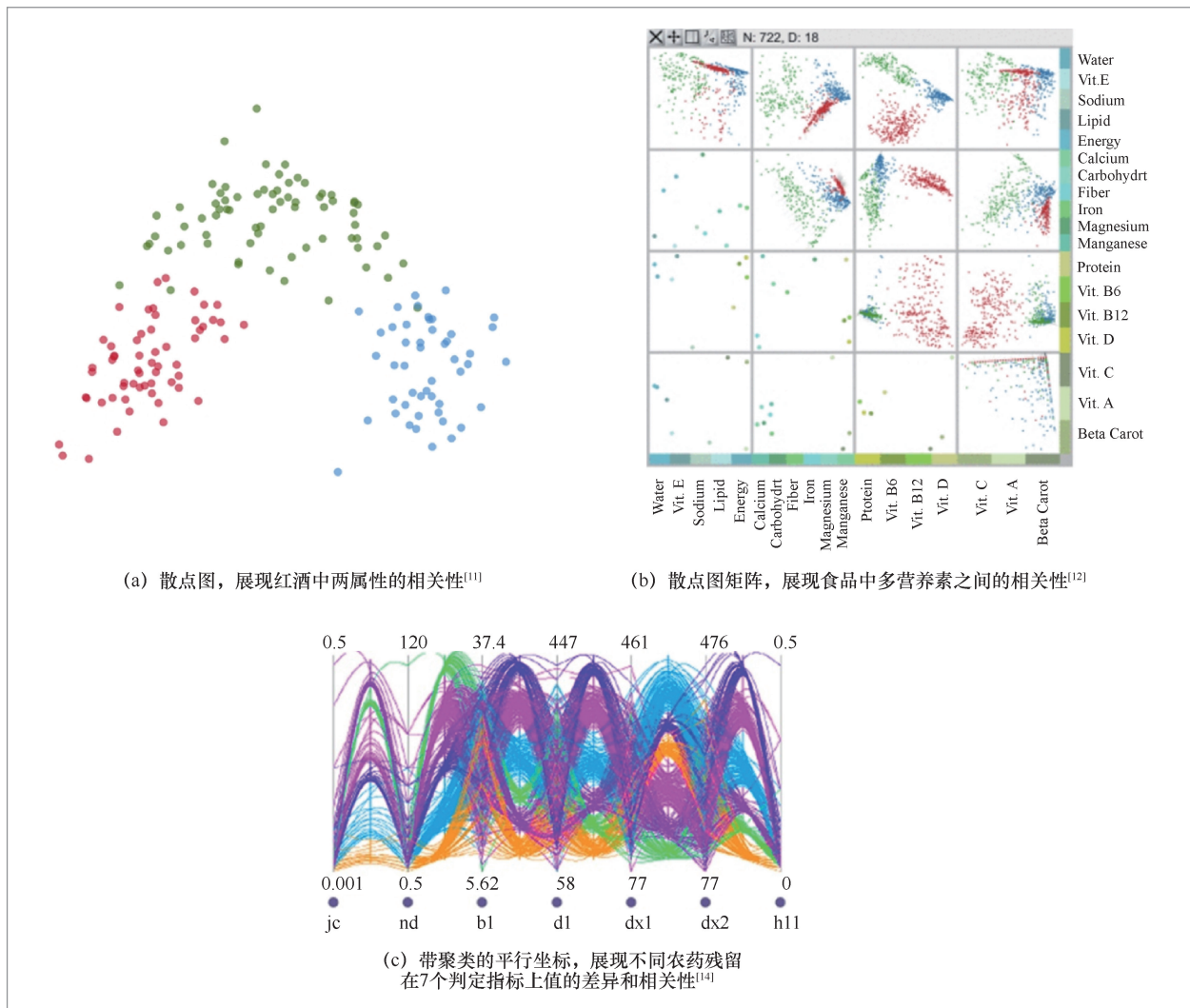


图1 属性关联可视分析方法图例

线<sup>[32-33]</sup>。采用平行坐标展现多属性数据可以帮助用户分析属性之间的相关性，以及数据的分布特征。Chen Y等人<sup>[13]</sup>运用平行坐标展现农产品中农药残留限量值在中国、美国、日本、欧盟、国际食品法典委员会(Codex Alimentarius Commission, CAC)等国家/地区或组织的MRL标准中的取值，帮助用户对比多个MRL标准的差异。平行坐标可以很好地展现属性间的相关关系，但当数据对象增多时，会产生视觉杂乱。为了解决这一问题，陈谊等人<sup>[14]</sup>对平行坐标进行了改进，提出了一种基于类

区间的多维数据可视化方法，并应用到农药残留检测标准数据的分析中，如图1(c)所示。该方法首先根据数据的实际理化意义，使用K-means算法对原始数据进行聚类，然后引入类区间模型，将各类数据相互分离，从而形成了清晰的可视化聚类效果。此外，陈谊等人<sup>[15]</sup>使用邻接矩阵，通过矩阵热图的颜色深浅映射维度之间的相关性大小，颜色越深表示相关性越大，并以检出农药为例，展示了不同农药属性维度之间的相关性，以挖掘某地区农产品中农药的施用模式。

### 3.2 实体间的关联可视分析

针对分析任务, 实体间的关联关系可以分为实体间网络关系和实体间层次关系。

#### 3.2.1 实体间网络关系可视分析

网络关系通常用图来描述, 在图结构中, 常将节点称为顶点, 边为顶点的有序偶对, 若两顶点间存在一条边, 则表示这两个顶点具有相邻关系<sup>[34]</sup>。基于图的可视分析方法<sup>[35-36]</sup>结合了图分析理论和可视化技术的优势, 将图的邻接表或者邻接矩阵转换为由点和线组成的图形, 可以令研究人员直观地看到数据间的关联关系, 其已成为分析复杂数据集中各种关系的有效手段。常用方法有节点-链接 (node-link)

法、邻接矩阵和弦图等。

节点-链接法是一种典型的关联数据可视化方法, 它用不同形状的节点表示实体, 节点之间的连线表示实体间的关系。它既可以表达实体间的树状层次关系 (节点链接树), 也可以表达实体间网状的关联关系 (图)。杨璐等人<sup>[16]</sup>运用节点-链接法将某种食品与其检出的不合格项目进行连线, 检出的不合格项目频率越高, 连线越粗, 与检测频率呈等比例关系展示, 从而帮助监管者定位重点监管对象和监管项目。弦图可以展示关联数据间的权重关系, 数据点之间的关系被绘制为连接两个数据点的弧 (边), 权重越大则边越粗。Narcisa P A 等人<sup>[17]</sup>设计了一种弦图来显示各危害物与乳制品之间的关联关系, 如图2所示。

邻接矩阵是实体间网络关系可视化表

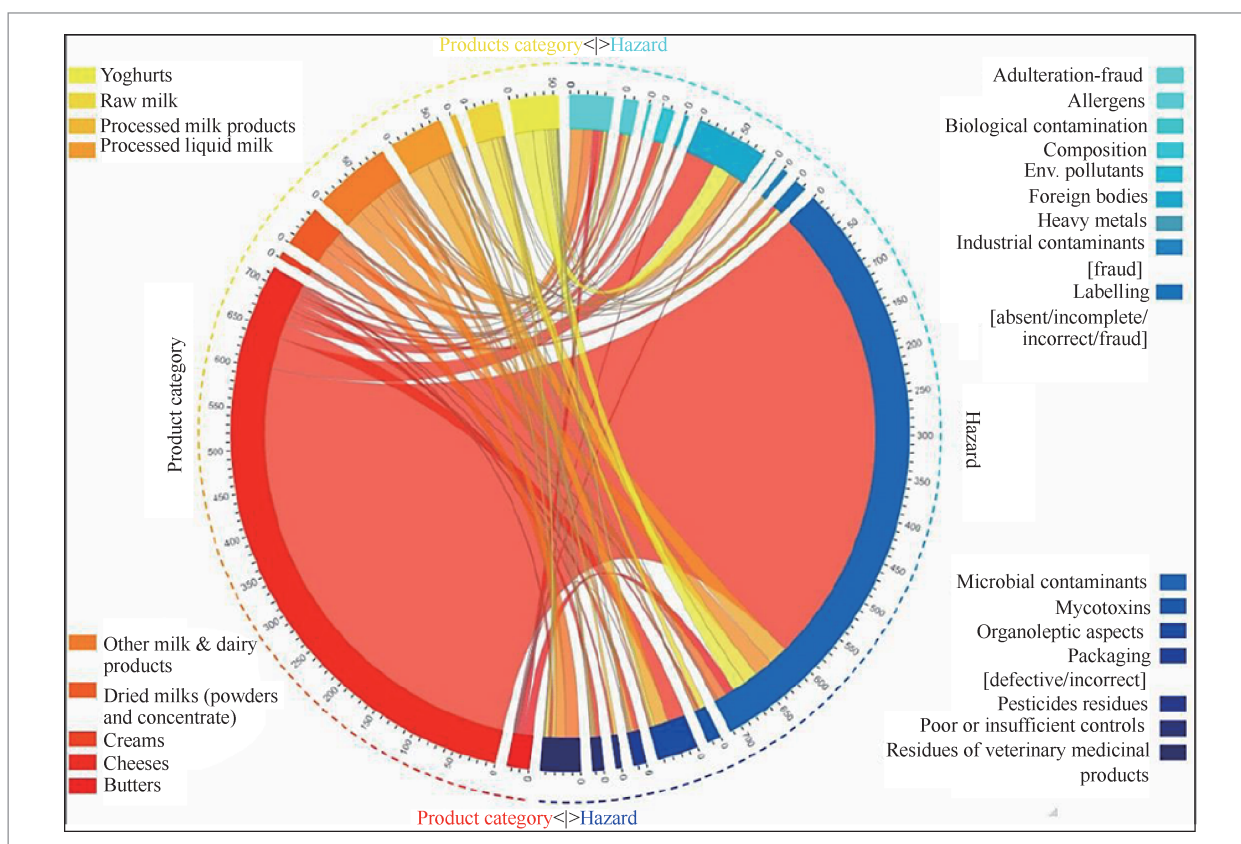


图2 弦图, 展现各危害物与乳制品之间的关联关系<sup>[17]</sup>

示之一,它是一个 $N \times N$ 的网格(其中 $N$ 为节点数),其中位置 $(i, j)$ 表示节点 $i$ 和 $j$ 之间的链路权值。Chen Y等人<sup>[18]</sup>提出了一种用于关联数据可视分析的有序矩阵表达方法,该方法用矩阵热图表示农产品和农药的检出关系,其中矩阵的行表示农药,列表示农产品,单元格的颜色表示农药残留含量,颜色越深,农药残余含量越高。如图3所示,该方法将数据抽象为二部图,将实体抽象为点, $P$ 、 $R$ 、 $W$ 分别表示两个独立的点集合和一个连接两点的边集合,将实

体的RW值作为索引,表示实体在关系结构中的重要性。RW-Rank算法受到Page-Rank算法的启发,根据行向量和列向量的RW值进行排序,创建一个有序的关系矩阵,帮助用户定位关键实体并分析它们之间的关系。如图4所示, $A_1$ 视图中矩阵 $A$ 的行和列的排列顺序是按字母顺序排序的,便于用户按名称查找; $A_2$ 视图中矩阵 $B$ 的行和列则按RW-Rank算法排序,便于用户快速定位高残留量、污染严重的农产品和农药。

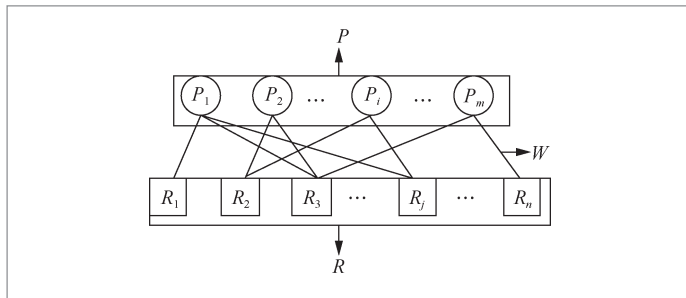


图3 二部图,展现农药和农产品的关系<sup>[18]</sup>

### 3.2.2 实体间层次关系可视分析

实体间层次关系主要表现为包含关系和从属关系。例如食品通常有大类、亚类、次亚类、细类之分。对于层次数据,常用的可视化方法有节点-链接树、树图、放射环等。

节点-链接树的层次关系表现为数据中父辈节点将子孙辈节点包围起来,展示

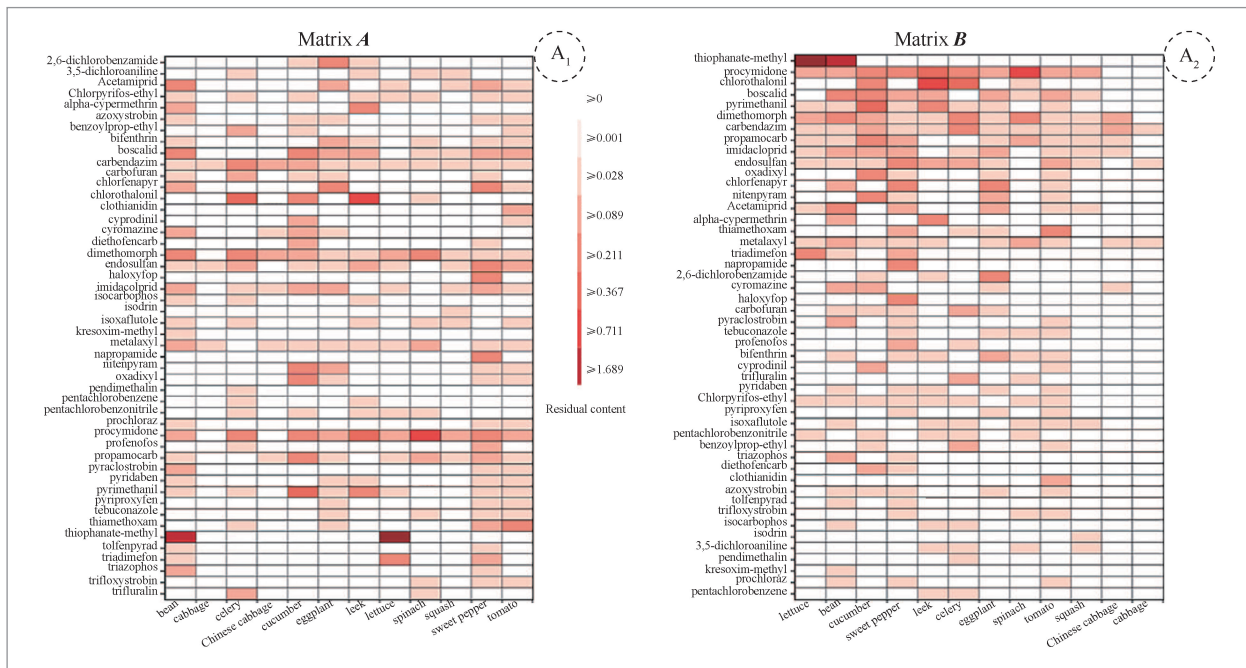


图4 邻接矩阵,展现农产品和农药的相关性<sup>[18]</sup>

了数据间的父子关系。齐红革等人<sup>[19]</sup>运用节点-链接树构建食品分类图谱展示食品分类,使食品、食品添加剂分类由大类到细类清晰地呈现。树图<sup>[37]</sup>由一系列嵌套的矩形组成,也可以说是在矩形空间中进行递归的分割,这些矩形的大小与相应的节点属性值成比例。大矩形表示数据树的一个分支,然后将大矩形细分为较小的矩形,表示该分支内每个节点的大小,还可以搭配颜色来表示不同的数据属性。Jia Y J等人<sup>[20]</sup>提出了一种基于树图的可视化方法,用树图表示农药含量的分布以及农产品所在区域的层次化结构,该方法可有效地帮助专家按照区域和农产品类别的层次结构对数据集进行关联分析。事实证明,利用树图的可视化方案能够有效地展示食品安全领域中对于层次结构和关联关系展示的要求。Chen Y等人<sup>[21]</sup>用树图保留节点间的层次关系,进而展示农药残留检测结果数据。图5(a)是天津市10个区2014年1月的果蔬农药残留分布情况,10个大矩形表示

天津市的10个区,每个大矩形中包含的两个小矩形分别表示水果和蔬菜,颜色表示农药残留超标率,即检测出的农药残留量超过MRL标准值的次数与检测农药残留总次数的比值,超标率越大越接近红色。通过这种方法,食品安全领域的专家可以更有效地发现各区果蔬农药残留的分布情况,并根据不同的需求做出有效的决策。

放射环也是一种展现层次关系的关联可视化方法,但是其内部空间利用率较低,因此,采用放射环和节点-链接法相结合的方式能够同时显示两类层次数据。Chen Y等人<sup>[22]</sup>充分利用放射环圆心附近的空间,将放射环与节点-链接树结合,提出了一种能够展示两类层次数据关联关系的可视化算法SONHC (sunburst with ordered nodes based on hierarchical clustering),并将其应用于可视分析农产品中农药残留的检出情况。该算法用外面的放射环展示农药的层次结构,圆环内部则通过节点-链接树展示农产品的层次结

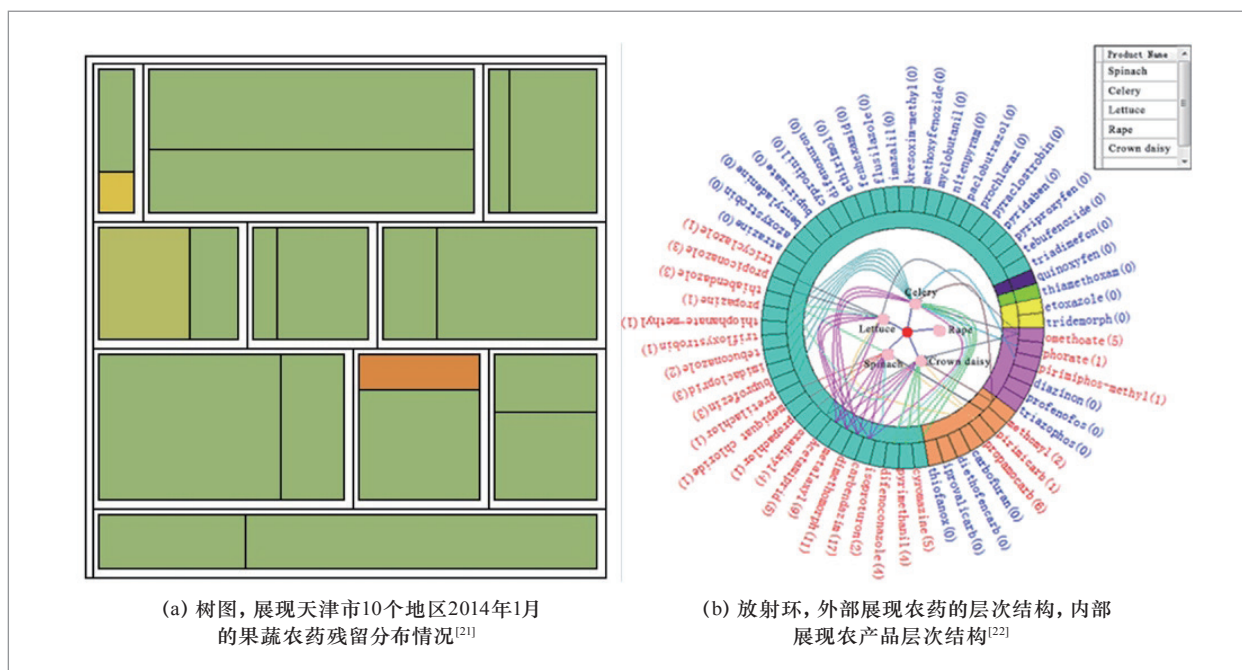


图5 实体间层次关系可视分析方法示例

构；通过连线将农产品与检出农药建立关联，帮助相关分析人员检查农药和农产品的层次结构，并探索农药和农产品之间的关联以及不同农药之间的关联，如图5(b)所示。此外，杜晓敏等人<sup>[23]</sup>基于圆环和放射环等提出了一种基于变换的可视分析关联图TransGraph，从而展示农产品与农药的检出关联关系和农药残留检出的层次关联关系。TransGraph能突出重点监管对象，全面展示关联信息，帮助相关监管部门和分析人员制定决策。

### 3.3 对比分析

对比分析是关联分析的一种常用方法，可分为数值对比和结构对比。

#### 3.3.1 数值对比可视分析

在食品安全领域中，常用数值对比来对比分析检测值与检测标准值，或将食品受污染情况进行排序对比。图6是一个多属性排名可视分析系统<sup>[24]</sup>，该系统采用平行坐标结合柱状图的可视化方法，展示多个农产品按多个农药残留评价指标（包括单因子和多因子综合指标）的排名情况，通过多视图联动、数据筛选、属性选择等交

互手段，帮助用户全面理解各种农产品受农药残留污染的排名情况，可对比分析各农产品及其农药残留情况。

#### 3.3.2 结构对比可视分析

在食品安全领域，通常需要比较两个按某种结构组织的数据集，如两个农产品分类体系的比较、两个MRL标准的比较都可以抽象为树比较的问题。结构对比可视分析最常用的方法是并置法，即将比较的数据直接可视化并放在同一个视图中，通过观察对比的方式完成<sup>[38]</sup>。Chen Y等人<sup>[13]</sup>将不同MRL限量标准抽象化为两个树，通过两个嵌套圆并置（如图7所示）来可视化两棵MRL树，帮助用户对比分析中国内地和中国香港的MRL标准。嵌套圆结合了节点-链接图和树图的优点，用圆的面积表示节点的属性值，用圆的嵌套关系表示节点间的层次关系，所有的子节点圆都被包含在父节点圆中。由于MRL标准是按农产品分类（树结构）来制定限量值的，该方法的两个MRL标准比较问题就转化为了两棵树的比较问题。用嵌套的结构描述农产品分类的层次结构，用圆的面积表示某农产品涉及MRL标准值的记录数，图7中的左嵌套圆为中国内地MRL标准，右嵌套圆为中国

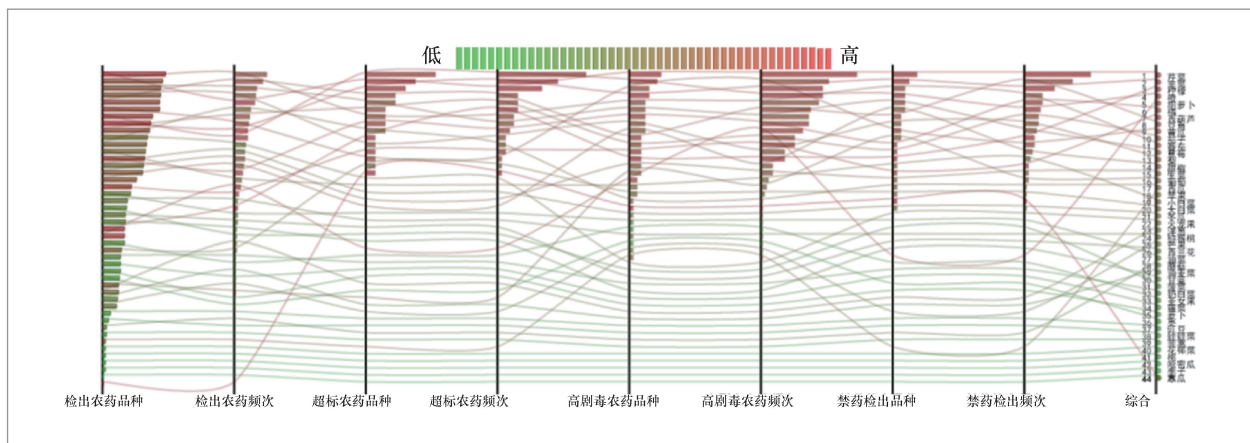


图6 多属性排名可视分析系统，展现多个农产品在多个农药残留评价指标的排名情况<sup>[24]</sup>

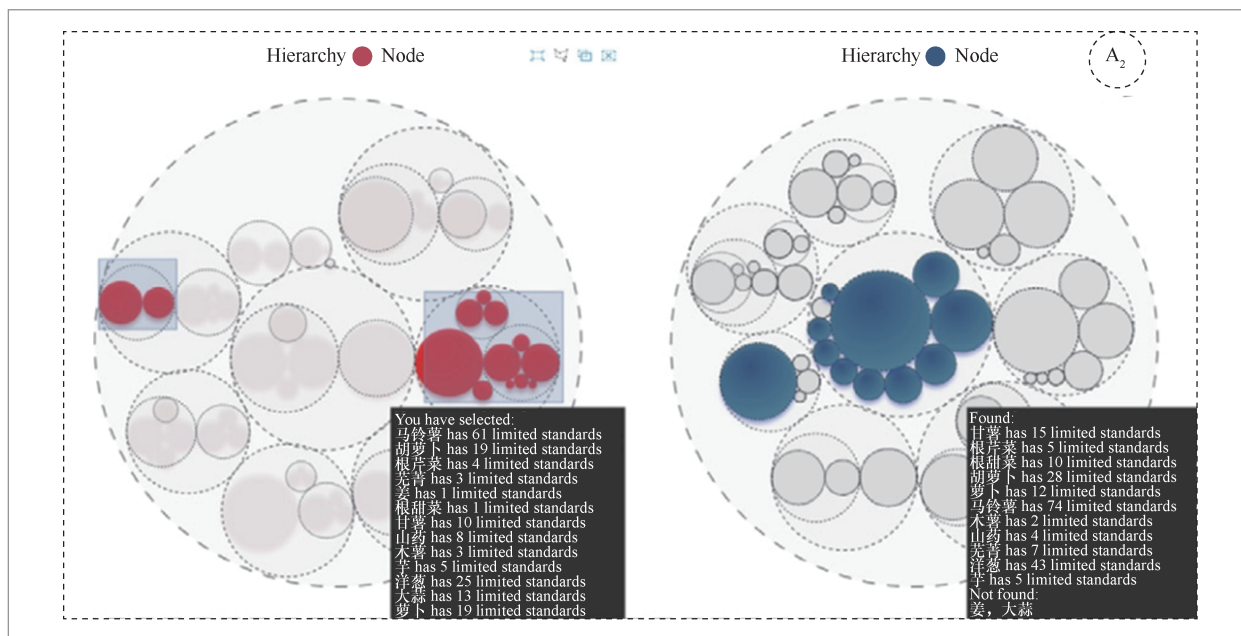


图7 嵌套圆，对比分析中国内地和中国香港的MRL标准<sup>[3]</sup>

香港MRL标准。

### 3.4 时空分析

食品数据具有时间和空间特征，对食品数据进行时空分析可以探索食品属性与时间、空间的关联，让分析人员掌握食品安全问题在地域上的分布特征和随时间推移的发展趋势。

#### 3.4.1 食品属性与时间的关联可视分析

食品安全数据在时间序列上存在的潜在规律（如多年农产品中的农药残留检出频次具有周期性特征，一年中不同季节的农药残留也呈现出不同的特征等）能有效地帮助人们对食品安全事件发生的时间进行预测，对安全问题进行提前预防。基于时间的可视化方法能够较为直观地反映数据随时间变化的规律和趋势，同时能够展现数据细节。时序型数据通常以时间线的方式来表示，图8(a)以时间为轴，展示

了7种农产品的农药超标率随时间的变化情况，其中气泡颜色表示农产品类型，大小表示农药残留超标率的值<sup>[25]</sup>。Chen Y等人<sup>[21]</sup>提出了一种被称为有序树图序列(ordered small multiple treemaps, OSMT)的时变层次数据可视化方法，它采用树图并置的方式实现对层次数据随时间变化的可视化表示。图8(b)为使用该方法对2014年天津市10个地区的果蔬农药残留超标率的变化情况进行可视化的结果，可以看出，图8(b)中的12个树图并置表示12个月的变化情况，每个展示层次数据的树图表示某个月天津市10个地区果蔬农药残留的超标率。主题河流(ThemeRiver)<sup>[39]</sup>也是一种常用的时间关联可视分析方法，其在时间维度上以河流的形式展现数据的变化情况。甄远刚等人<sup>[26]</sup>提出了一种改进的非连续数据ThemeRiver可视化方法，利用高斯模型曲线拟合，通过主题布局排序、颜色选择和标签分布布局一种具有预测功能且能够展示层次特征的新型主题河流模型，并将其应用于农药残留数据监测，为农

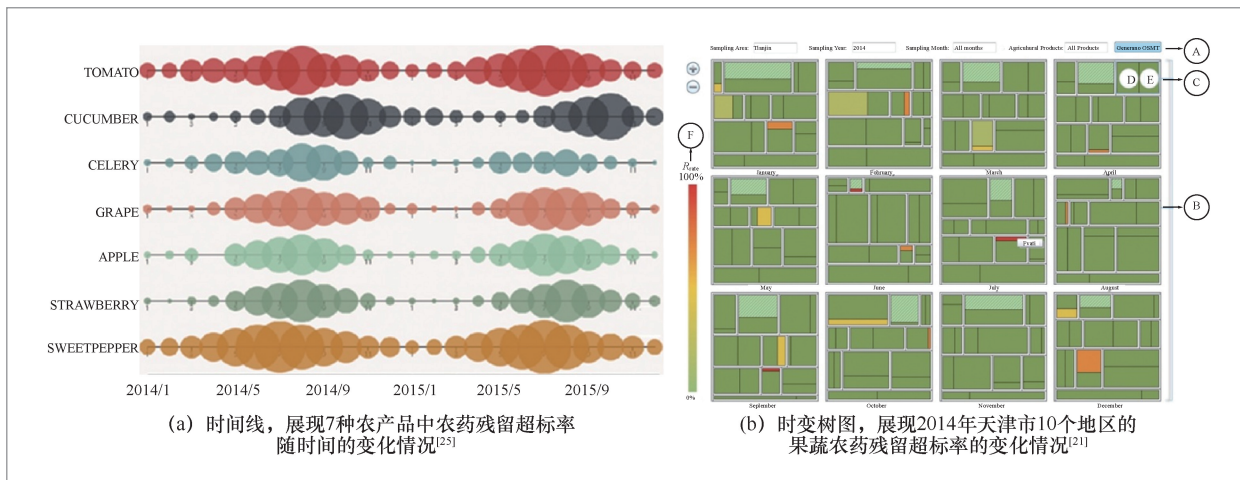


图8 时间关联可视分析方法示例

药残留预测预警提供依据。

### 3.4.2 食品属性与空间的关联可视分析

食品安全数据的地域分布能够帮助研究人员定位食品的采样地、原产地等，是食品安全预警和溯源的重要依据。通常地理信息系统 (geographic information system, GIS) 被用来显示数据在地域上的分布情况。统计地图 (choropleth map) 可视化假设数据的属性在一个区域内部平均分布，通过颜色表示数据的内在模式。庞国芳等人<sup>[27]</sup>研发了农药残留可视化在线制图系统，编制了中国市售水果蔬菜农药残留水平地图集，展示了不同MRL标准下的农药超标情况。

统计地图最大的问题在于数据分布和地理区域大小的不对称性。变形地图 (cartogram) 可以解决此问题，变形地图可以依据某个专题属性对地理要素进行扭曲、变形<sup>[40]</sup>，用相对属性值的大小取代真实面积，利用夸张的效果更直观地反映数量特征。Vanasse A等人<sup>[28]</sup>比较了加拿大各地区的肥胖率，并评估了各地区的成人肥胖率、休闲时间体育活动水平和水果蔬

菜摄入量之间的生态关系。由于各地区人口密度不同，为了更科学地反映结果，将地图按人口密度进行了调整，如图9(a)所示。Plaza-Rodríguez C等人<sup>[29]</sup>应用变形地图探讨德国各州零售生鸡肉样品中弯曲杆菌检出率的地区分布，如图9(b)所示，德国各州的大小根据弯曲杆菌检出率进行了修改，导致地图的原始地理形状和拓扑结构被扭曲了，让地理面积小且检出率高的地区也同样能受到关注，食用过此鸡肉的患者也能被及时监管。Chen Y等人<sup>[25]</sup>创建了中国农药残留分布地图，统计地图中渐变的紫色区域表示抽样农产品中检测到的农药残留量超过MRL标准值的频次，变形地图使用扩散算法基于上述频率值生成。

## 4 问题与挑战

从前文可以看出，大数据技术已成为食品安全风险监控的重要手段，可视分析在食品安全风险分析和预警方面开始发挥越来越重要的作用，并已取得一定进展，但仍面临一些问题和挑战。

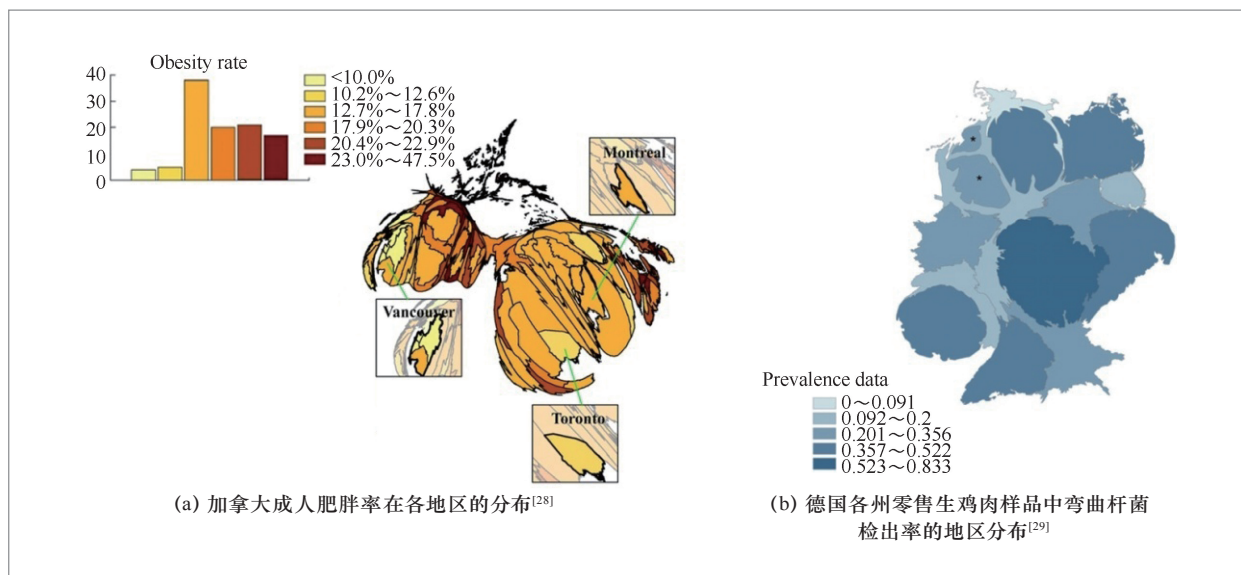


图9 空间关联分析可视分析方法图例

### (1) 多源异构数据的融合处理与可视分析

随着检测技术的进步和计算机网络技术的普及、食品安全数据的规模快速增长,数据的准备和处理呈现出人力成本高和时间周期长两大特点<sup>[41]</sup>,数据类型也呈现出多样化特点,包括数值、文本、图像、视频或这些数据类型的组合。近年来,针对数值型结构化数据的可视分析技术和方法已取得了丰富的成果,但针对文本、图像、视频等非结构化数据的分析仍面临挑战。如何在可视分析流程中结合数据分析、文本挖掘、图像识别、视频处理等技术实现对多模态数据的可视分析,帮助食品安全领域的人员实现风险识别、风险发现、风险分析,仍是一个亟待解决的问题。

### (2) 人工智能在可视分析过程中的应用

以机器学习为代表的人工智能技术已开始应用于可视分析流程中的各个环节。在数据分析阶段,人工智能可以为数据的预处理、数据变换、数据投影等多个步

骤提供准确高效的方法,也可以为食品安全风险评估、趋势预测提供科学的模型;在可视化阶段,应用人工智能技术对数据和分析任务进行处理、分析和学习,实现对可视化设计方案的选择,包括映射方式(散点图、平行坐标、节点-链接等)、布局(正交、径向等)、配色方案等,自动生成可视化解决方案<sup>[42]</sup>,从而减轻领域人员在可视化设计方面的负担。

### (3) 设计易于理解的可视分析系统

从事食品安全数据分析工作的人员通常不是计算机领域专业人员,现有大部分可视分析系统提供的可视化表达相对复杂,对于领域用户,即使其具有丰富的领域专业知识和经验,也需要经过一定程度的培训后才能熟练使用,这就限制了可视分析系统的广泛应用。因此,设计更易于理解和使用的可视化表达(如对人们熟悉的散点图、柱状图、节点-链接等进行改进)、开发方便易用的交互方式(如应用虚拟现实和增强现实设备实现更自然的人与数据的交互),也是当前亟待解决的问题。

## 5 结束语

食品安全大数据可视分析技术通过提供图形化的交互界面,帮助用户洞悉蕴含在数据中的现象和规律,提高对食品安全风险的分析、发现、预警和溯源能力,为食品安全监测和管控提供了新手段。本文分析了食品安全大数据的主要来源、特征和分析任务,提出了一种关联可视分析技术分类方法,从属性关联、实体关联、对比分析和时空分析4个方面总结了食品安全大数据可视化关联分析方法,最后从多源异构数据的融合处理与可视分析、人工智能在可视分析过程中的应用、设计易于理解的可视分析系统3个方面提出了该领域存在的问题、机遇和挑战。未来工作是将人工智能与可视分析技术结合,特别是使用结合数据挖掘、机器学习、深度学习的可视分析方法,解决食品安全风险分析、风险识别、风险预警和风险溯源等问题,提高对食品安全的监测和管控能力。

## 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.  
XU R G, PANG G F. A strategic study on the current situation, problems and countermeasures of food safety in China[M]. Beijing: Beijing Science Press, 2015.
- [2] 陈谊, 刘莹, 田帅, 等. 食品安全大数据可视分析方法研究[J]. 计算机辅助设计与图形学学报, 2017, 29(1): 8-16.  
CHEN Y, LIU Y, TIAN S, et al. A survey of visual analytical techniques for big data in food safety field[J]. Journal of Computer-Aided Design & Computer Graphics, 2017, 29(1): 8-16.
- [3] MARVIN H J P, JANSSEN E M, BOUZEMBRAK Y, et al. Big data in food safety: an overview[J]. Critical Reviews in Food Science and Nutrition, 2016, 57(11): 2286-2295.
- [4] JIN C Y, BOUZEMBRAK Y, ZHOU J H, et al. Big data in food safety—a review[J]. Current Opinion in Food Science, 2020, 36: 24-32.
- [5] MAURO A, GRECO M, GRIMALDI M. What is big data? A consensual definition and a review of key research topics[C]// The 4 th International Conference of Integrated Information. [S.l.:s.n.], 2015: 97-104.
- [6] WU Y N, CHEN J S. Food safety monitoring and surveillance in China: past, present and future[J]. Food Control, 2018, 90: 429-439.
- [7] HU K, LIU J, LI B, et al. Global research trends in food safety in agriculture and industry from 1991 to 2018: a data-driven analysis[J]. Trends in Food Science & Technology, 2019, 85: 262-276.
- [8] 许世卫. 农业大数据与农产品监测预警[J]. 中国农业科技导报, 2014, 16(5): 14-20.  
XU S W. Agricultural big data and monitoring and early warning of agricultural products[J]. Journal of Agricultural Science and Technology, 2014, 16(5): 14-20.
- [9] 张平文, 鄂维南, 袁晓如, 等. 大数据分析与应用技术创新平台[J]. 大数据, 2018, 4(4): 85-95.  
ZHANG P W, E W N, YUAN X R, et al. Big data analysis and application technology innovation platform[J]. Big Data Research, 2018, 4(4): 85-95.
- [10] MUNZNER T. Visualization analysis and design[M]. Boca Raton: CRC Press, 2014.
- [11] BIAN R Z, XUE Y M, ZHOU L, et al. Implicit multidimensional projection of local subspaces[J]. IEEE Transactions on Visualization and Computer Graphics, 2020, 27(2): 1558-1568.

- [12] YUAN X R, REN D H, WANG Z C, et al. Dimension projection matrix/tree: interactive subspace visual exploration and analysis of high dimensional data[J]. IEEE Transactions on Visualization and Computer Graphics, 2013, 19(12): 2625–2633.
- [13] CHEN Y, DONG Y, SUN Y H, et al. A multi-comparable visual analytic approach for complex hierarchical data[J]. Journal of Visual Languages and Computing, 2018, 47: 19–30.
- [14] 陈谊, 李潇潇, 蔡进峰, 等. 基于类区间的多维数据可视化方法[J]. 系统仿真学报, 2013(10): 2418–2423.  
CHEN Y, LI X X, CAI J F, et al. Visualization of multi-dimensional data based on cluster region[J]. Journal of System Simulation, 2013(10): 2418–2423.
- [15] 陈谊, 张聪. 一种基于维度投影的多维数据相关性可视分析方法[J]. 计算机辅助设计与图形学学报, 2018, 30(4): 592–601.  
CHEN Y, ZHANG C. Visual analysis of correlation in multidimensional data based on dimension projection technique[J]. Journal of Computer-Aided Design & Computer Graphics, 2018, 30(4): 592–601.
- [16] 杨璐, 张馨月, 郑丽敏. 挖掘数据关系的食品抽检数据可视化分析图研究[J]. 农业机械学报, 2019, 50(6): 272–279.  
YANG L, ZHANG X Y, ZHENG L M. Visual analysis graph research of food sampling data based on mining data relationship[J]. Transactions of the Chinese Society of Agricultural Machinery, 2019, 50(6): 272–279.
- [17] NARCISA P A, SERGIU C S, MIRELA A A, et al. Analysis of RASFF notifications on contaminated dairy products from the last two decades: 2000–2020[J]. Romanian Biotechnological Letters, 2020, 25(2): 1396–1406.
- [18] CHEN Y, LYV C, LI Y, et al. Ordered matrix representation supporting the visual analysis of associated data[J]. Science China Information Sciences, 2020, 63.
- [19] 齐红革, 谭亚军, 黄琳琳, 等. 食品安全数据分析可视化模型研究[J]. 食品安全质量检测学报, 2019, 10(17): 5968–5973.  
QI H G, TAN Y J, HUANG L L, et al. Research on visualization model of food safety data analysis[J]. Journal of Food Safety and Quality, 2019, 10(17): 5968–5973.
- [20] JIA Y J, CHEN Y, LI Z G. Treemap-based visualization methods for pesticide residues detection data[C]// Advances in Image and Graphics Technologies. Heidelberg: Springer, 2013: 154–162.
- [21] CHEN Y, DU X M, YUAN X R. Ordered small multiple treemaps for visualizing time-varying hierarchical pesticide residue data[J]. Science China Information Sciences, 2017, 33(6–8): 1073–1084.
- [22] CHEN Y, ZHANG X Y, FENG Y C, et al. Sunburst with ordered nodes based on hierarchical clustering: a visual analyzing method for associated hierarchical pesticide residue data[J]. Journal of Visualization, 2015, 18(2): 237–254.
- [23] 杜晓敏, 陈谊, 李玥. TransGraph: 一种基于变换的可视分析关联图[J]. 计算机辅助设计与图形学学报, 2018(1): 79–89.  
DU X M, CHEN Y, LI Y. TransGraph: a transformation-based graph for analyzing relations in data set[J]. Journal of Computer-Aided Design & Computer Graphics, 2018(1): 79–89.
- [24] 陈谊, 刘莹, 陈星如, 等. 基于可视分析的农药残留污染仿真评估方法[J]. 计算机仿真, 2017, 34(10): 347–351.  
CHEN Y, LIU Y, CHEN X R, et al. Simulation and evaluation method for pesticide residue pollution based on visual analysis techniques[J]. Computer Simulation, 2017, 34(10): 347–351.
- [25] CHEN Y, ZHAO Y F, CHEN X R, et al. Visualizing geospatial distribution of pesticide residue pollution using cartogram and heat map[M]// Transactions on Edutainment XIII. Heidelberg: Springer, 2017: 231–244.

- [26] 甄远刚, 陈谊, 刘莹, 等. 一种基于 ThemeRiver模型的非连续层次数据可视化方法[J]. 系统仿真学报, 2015, 27(10): 2460–2466,2474.  
ZHEN Y G, CHEN Y, LIU Y, et al. Discontinuous hierarchical data visualization method based on ThemeRiver model[J]. Journal of System Simulation, 2015, 27(10): 2460–2466,2474.
- [27] 庞国芳, 等. 食用农产品农药残留监测与风险评估溯源技术研究[M]. 北京: 科学出版社, 2018.  
PANG G F, et al. Research on monitoring, risk assessment and traceability technology of pesticide residues in edible agricultural products[M]. Beijing: Science Press, 2018.
- [28] VANASSE A, DEMERS M, HEMIARI A, et al. Obesity in Canada: where and how many?[J]. International Journal of Obesity, 2006, 30(4): 677–683.
- [29] PLAZA-RODRÍGUEZ C, APPEL B, KAESBOHRER A, et al. Discussing state-of-the-art spatial visualization techniques applicable for the epidemiological surveillance data on the example of campylobacter spp. in raw chicken meat[J]. Zoonoses and Public Health, 2016, 63(5): 358–369.
- [30] CLEVELAND W S. Visualizing data[J]. Infection Control and Hospital Epidemiology, 1994, 15(12): 763–763.
- [31] CARR D B, LITTLEFIELD R J, NICHOLSON W L. Scatterplot matrix techniques for large N[J]. Journal of the American Statistical Association, 1987.
- [32] ZHOU L, WEISKOPF D. Indexed-points parallel coordinates visualization of multivariate correlations[J]. IEEE Transactions on Visualization and Computer Graphics, 2018, 24(6): 1997–2010.
- [33] FERDOSI B, ROERDINK J. Visualizing high-dimensional structures by dimension ordering and filtering using subspace analysis[J]. Computer Graphics Forum, 2011, 30(3): 1121–1130.
- [34] 陈为, 赵焯, 张嵩, 等. 可视化导论[M]. 北京: 高等教育出版社, 2020.  
CHEN W, ZHAO Y, ZHANG S, et al. Introduction to visualization[M]. Beijing: Higher Education Press, 2020.
- [35] CHEN Y, GUAN Z L, ZHANG R, et al. A survey on visualization approaches for exploring association relationships in graph data[J]. Journal of Visualization, 2019, 22(3): 625–639.
- [36] 程致远, 鲍玉斌, 冷芳玲. 面向大规模图数据的并行图布局算法[J]. 大数据, 2016, 2(5): 12–21.  
CHENG Z Y, BAO Y B, LENG F L. Parallel graph layout algorithm for large-scale graph data[J]. Big Data Research, 2016, 2(5): 12–21.
- [37] 张昕, 袁晓如. 树图可视化[J]. 计算机辅助设计与图形学学报, 2012, 24(9): 1113–1124.  
ZHANG X, YUAN X R. Treemap visualization[J]. Journal of Computer-Aided Design & Computer Graphics, 2012, 24(9): 1113–1124.
- [38] 李彦龙, 李国强, 董笑菊. 树比较可视化方法综述[J]. 软件学报, 2016, 27(5): 1074–1090.  
LI Y L, LI G Q, DONG X J. Survey on visualization of tree comparison[J]. Journal of Software, 2016, 27(5): 1074–1090.
- [39] HAVRE S, HETZLER E G, WHITNEY P. ThemeRiver: visualizing thematic changes in large document collections[J]. IEEE Transactions on Visualization & Computer Graphics, 2002, 8(1): 9–20.
- [40] GASTNER M T, NEWMAN M E J. Diffusion-based method for producing density-equalizing maps[J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(20): 7499–7504.
- [41] 范举, 陈跃国, 杜小勇. 人在回路的数据准备技术研究进展[J]. 大数据, 2019, 5(6): 3–18.  
FAN J, CHEN Y G, DU X Y. Progress on human-in-the-loop data preparation[J]. Big Data Research, 2019, 5(6): 3–18.

[42] WANG Y H, HAN F B, ZHU L F, et al.  
Line graph or scatter plot? Automatic  
selection of methods for visualizing trends

in time series[J]. IEEE Transactions on  
Visualization and Computer Graphics,  
2018, 24(2): 1141-1154.

#### 作者简介



陈谊(1963- ),女,博士,北京工商大学教授,食品安全大数据技术北京市重点实验室主任,中国图象图形学学会可视化与可视分析专业委员会副主任,中国计算机学会杰出会员。主要研究方向为可视化与可视分析、智能信息处理、食品安全大数据技术。2016年获得中国分析测试协会科学技术奖(CAIA奖)特等奖,2017年获得中国石油和化工科技进步奖二等奖。多次担任PacificVis、ChinaVis、ChinaVR等可视化学术会议程序委员会委员和审稿人。



孙梦(1996- ),女,北京工商大学计算机学院硕士生,主要研究方向为可视化与可视分析、食品安全大数据技术。



武彩霞(1998- ),女,北京工商大学计算机学院硕士生,主要研究方向为可视化与可视分析。



孙小然(1997- ),女,北京工商大学计算机学院硕士生,主要研究方向为可视化与可视分析。

收稿日期: 2021-01-07

通信作者: 陈谊, chenyl@btbu.edu.cn

基金项目: 国家重点研发计划资助项目(No.2018YFC1603602); 国家自然科学基金资助项目(No.61972010)

Foundation Items: The National Key Research and Development Program of China(No.2018YFC1603602), The National Natural Science Foundation of China(No.61972010)