

doi:10.6041/j.issn.1000-1298.2019.06.031

挖掘数据关系的食品抽检数据可视化分析图研究

杨璐^{1,2} 张馨月¹ 郑丽敏^{1,2}

(1. 中国农业大学信息与电气工程学院, 北京 100083; 2. 食品质量与安全北京实验室, 北京 100083)

摘要: 食品安全数据具有多源、关联和不确定性等特征,数据的项目、项目的属性以及相链接数目较多,数据内部潜在关系不明晰,需要研究能够进行关系挖掘的可视视图。针对食品安全领域数据分析的实际需求,采用圆环布局、节点链接布局等元素,对数据间的简单关系和层次结构进行展示;结合同心圆布局、散点图、热力图元素和动态过滤以及数据聚类技术,在展示数据节点性质的同时,揭示数据间的潜在关联关系,并综合以上视图提出了一种挖掘数据关系的可视分析图 ExploreView。应用于国家食品药品监督管理总局抽检数据集,使用立方体隐喻组织数据,二分图定义任务需求,完成可视编码,进行数据关系探索,为可能发生的食品安全事件提供预警,定位重点监管对象,为制定新的食品安全规章制度提供参考。

关键词: 食品抽检; 数据关系挖掘; 可视化

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 1000-1298(2019)06-0272-08

Visual Analysis Graph Research of Food Sampling Data Based on Mining Data Relationship

YANG Lu^{1,2} ZHANG Xinyue¹ ZHENG Limin^{1,2}

(1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

2. Beijing Laboratory of Food Quality and Safety, Beijing 100083, China)

Abstract: The relationship between data can be visualized by using multiple types of images, so it is convenient for users to obtain information and relationships between data. However, when many data items, attributes and links, and the relationships between data are not clear, a visual view which is capable of relationship mining is required. For the real task requirements of domain data analysis, elements such as circle layout and node link layout were used to display simple relationships and hierarchical structures between data. What's more, combining concentric circle layout, scatter plots, thermogram elements and dynamic filtering and data clustering techniques, a relational mining view was proposed that not only demonstrated the nature of data nodes, but also revealed the potential relationships between data. Finally, combining the above views, a visual analysis graph of the mining data relationship was presented, which was ExploreView. It was applied to the sampling data set of the Food and Drug Administration, while using cube metaphor to organize data. The bipartite graph defined task requirements before completing visual coding. It can display the basic situation of data information and dynamically interact according to the actual needs of users, and reflect the attributes, various hierarchical structures and relationships between data. As a result, the visual analysis graph was easy and efficient to operate. It can be used to provide early warning for possible food safety incidents, locate key regulatory targets, which provided reference for the development of rules, and effectively met the needs of different types of users.

Key words: food sampling; data relationship mining; visualization

0 引言

可视化是把数据信息转换为人眼能识别的可视

符号,是人机交互、虚拟现实应用的关键技术^[1]。图可视化是信息可视化领域的一个重要分支,又称为关联网络、网络或关联图,是关联数据分析过程中

收稿日期: 2019-03-11 修回日期: 2019-04-18

基金项目: 国家重点研发计划项目(2017YFC1601803)和国家蛋鸡产业技术体系项目(CARS-40-K27)

作者简介: 杨璐(1967—),女,副教授,博士生,主要从事人工智能与网络工程研究,E-mail: yanglumail@263.net

的重要部分^[2]。

为了对结构语义进行可视化,研究者提出了文本的语义结构可视化技术^[3]。DAViewer将文本的叙述结构语义以树的形式进行可视化,同时进行了相似度统计、修辞结构,以及展示了相应的文本内容^[4]。DocuBurst以放射状层次圆环的形式表示文本结构^[5]。基于主题文本聚类是文本数据挖掘的重要研究内容,为了可视化表示文本聚类效果,通常将一维的文本信息投射到二维空间中,以便于对聚类的关系予以展示。Hipp是一种基于层次化点排布的投影方法,可广泛用于文本聚类可视化^[6]。在面对高维数据时,LAI等^[7]提出一种通过局部增强的投影探索高维数据的交互式探索方案,以帮助用户根据他们的兴趣点和分析任务定制线性投影。

可视化分析视图或分析系统是一种辅助数据分析和决策的技术,能够让用户更好地理解复杂的领域知识。CHEN等^[8]根据农药残留数据的特点和食品安全领域的需求,提出一种基于层次聚类的关联层次数据可视化分析方法。CHEN等^[9]在面临综合比较多区域MRL标准数据的挑战时,提出一种复杂层次数据多重比较可视化分析方法。CHUI等^[10]以多面板图展现年龄、时间和疾病之间的相互作用,帮助流行病学家构造了视觉分析图。LI等^[11]提出了为不同背景的房地产市场用户提供服务、满足不同程度需求的交互式视觉分析系统。

近年来,农产品及食品质量安全事件时有发生,在给消费者带来健康威胁的同时,也给行业带来了冲击^[12]。食品安全相关的大数据通常具有多源、多维、时空、层次、关联和不确定性等特征^[13]。而对于食品安全这一特殊领域,监控、溯源食品安全事件,进行风险评估和预警是十分重要的^[14]。食品安全检测是食品安全监管的重要环节,检测环节的薄弱和缺失是我国出现食品安全问题的直接原因^[15-16]。借助于可视化方式,能够清晰地展现数据的表面特征,挖掘数据的内在关系,总结不合格产品的产生原因,对评价当前食品安全形势、查找风险因素、引导监管方向、推动食品行业健康发展具有重要作用。

为更好地挖掘以食品抽检不合格报告为代表的关系,本文提出一种数据关系挖掘可视分析图——ExploreView,展现数据集中的层次结构,同时提供隐藏在数据层次结构中的关联关系。

1 研究对象

以2017年12月25日至2018年1月15日间的国家食品药品监督管理总局抽检不合格报告的数据集为研究对象,共包含1 014条抽检数据。不合

格抽检报告数据集一般包括标称生产企业名称、企业地址、被抽样单位名称、单位地址、食品名称、规格型号、商标、生产日期及批号、不合格项目名称、检验结果、标准值、食品分类、公告号、公告日期、任务来源和检验机构等内容。

对数据集进行数据预处理、规范化和特征提取;根据数据集的特点,建立数据模型,分析实际任务需求并进行映射;最后进行视觉元素编码,完成可视化。

2 可视化基本模型

2.1 数据立方体

数据立方体是多维模型的形象概念。虽然立方体只有三维,但多维模型不限于此,可以组合更多的维度,展示更多维度的数据特征。食品的抽检不合格报告维数众多,经过数据预处理、转换,根据实际需要,能够按照预先定义的多维模型,将数据加载进去。

实现多维数据操作的一种方法是使用维的层次关系,每个层次包含一个或多个维成员,多个层次中的维成员排列成多个层次,这种层次结构为数据的上钻和下钻提供了基础。实现多维数据操作的另一种方法是使用维内元素的类,对应类的属性对维成员进行划分,类属性不同,得到的划分也不同。层次和类是两个不同的概念,层次有父子关系,而类的成员之间不存在父子关系,只反映成员的共同特征。

在对食品抽检报告的分析中,既有按维的层次关系进行的分析,也有按维成员的类进行的分析。在层次关系上进行的分析,包括从维的低层次到高层次的数据综合,即聚集分析,以及从维的高层次到低层次的数据钻取分析。成员类的分析包括分类与归纳,分类是由粗粒度成员向细粒度成员进行的下探,归纳则是由细粒度成员向粗粒度成员进行的汇总。

聚合、组织数据集是将数据按照所需的 n 个维度进行转换。比如,选择食品、抽检项目和地点3个维度,分别表示为 F (食品)、 P (项目)和 L (地点),则笛卡尔乘积 $F \times P \times L$ 隐喻地称为数据立方体^[17]。出于实际目的,3个集合 F 、 P 和 L 离散且有限。由指定食品 $f \in F$ 、抽检项目 $p \in P$ 和地点 $l \in L$ 组成的任何组合称为立方体的点,表示为 (f, p, l) 。

2.2 切片和投影

任务分析是探索数据集三维空间中的研究对象的变化和关系。但是,不能从结构 $F \times P \times L$ 的高维度看到这种变化和关系,所以整个任务被分解成为更简单的子任务。可以通过考虑其切片来完成分析,其中一个维度固定,探究剩余维度变量的变化

和关系。例如,在 $F \times P \times L$ 所构成的数据立方体中,为了探究 F 和 P 之间的关系,就需要进行切片。在地点维上选定一个维成员,可得到地点维的一个切片。

将一些汇总运算符(求和、平均值等)应用在一个维变量的多个值的多个切片上,可以聚合这些切片,称为投影,如图 1b 所示。投影的数据结构和获得它的切片(图 1a)的数据结构相同,投影结构中的值是来自切片的值的聚合^[18]。

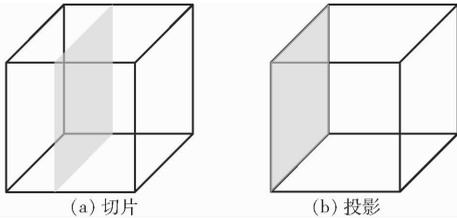


图 1 切片与投影模型示意图

Fig. 1 Schematic diagrams of slice and projection model

2.3 二分图

本文用三维数据立方体的方式来简化多维模型进行描述,并将其他维度作为切片或投影中的变量,使用二分图进行表示。

通常可以形式化将图表述为二元组 (V, E) , 即 $G = (V, E)$, 其中 V 为顶点集合, E 为顶点间关联关系所对应的边集合^[19]。二分图也称二部图,为图论中的一种特殊模型,该类图的特点是其顶点集可以分成 2 个不相交的集合,使得在同一个集合内的顶点互不相邻(没有共同边)^[20]。如果 $M \subseteq E$, 并且 M 中没有任何两边有公共端点且不含环,则称 M 为 G 的一个匹配。

食品和地点形成二分图,其中每个边表示某个地点某种食品抽检出不合格项目。类似地,食品和抽检项目也形成二分图,其中边表示某种食品抽检出某种不合格项目,如图 2 所示。

3 可视化设计

3.1 任务需求分析

食品抽检报告内容丰富,维数众多,可根据需要选取适当维度映射为多维模型。本文以数据立方体

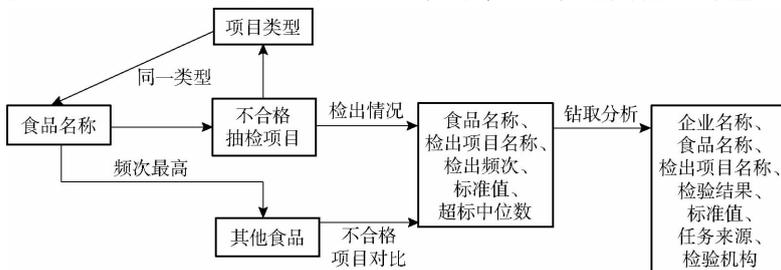


图 3 食品与抽检项目关系任务分析图

Fig. 3 Schematic of data cube and bipartite graph model mapping

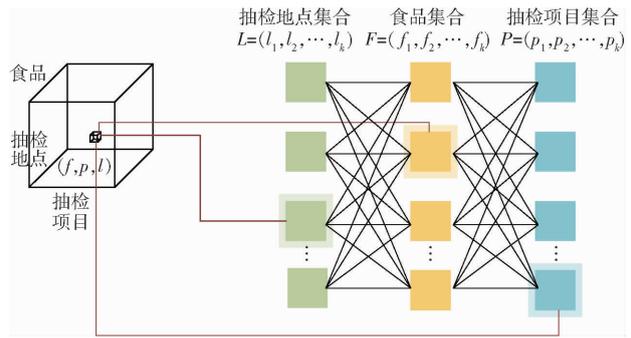


图 2 数据立方体与二分图模型映射示意图

Fig. 2 Schematic of data cube and bipartite graph model mapping

为例来进行分析。经过数据处理和特征提取,选择 3 个维度 F 、 P 和 L 。在 $F \times P \times L$ 数据立方体取一个切片,切点与其中一个维度内的一个选定值相对应,并通过二分图表示。相应的分析任务就在切片内进行,以降低多个变量变化造成的任务复杂度。基于数据立方体切片的任务包括:

(1) $l \rightarrow (F \times P \rightarrow M)$: 对于选定的抽检地点 $l \in L$, M 是食品集合 F 和抽检项目集合 P 形成的二分图的匹配。不同食品 and 不合格抽检项目的关系如图 3 所示。通过指定食品或样品名称筛选出不合格项目检出情况,包括食品名称、检出项目名称、检出频次、标准值、超标中位数等。进一步钻取相关抽检条目的具体信息,包括检验结果、检验机构等。筛选出与某些特定食品检出同一不合格项目的其他食品,并进行不合格项目的对比,包括项目名称、检验结果、标准值、检出频次等。展示检出不合格项目属于一定项目类别内的食品类型。分析检出不合格项目频次最高的食品种类名称,并对应显示检出项目的名称、频次、检验结果和标准值对比。

(2) $f \rightarrow (L \times P \rightarrow M)$: 对于选定的被抽检食品 $f \in F$, M 是抽检地点集合 L 和抽检项目集合 P 形成的二分图的匹配。研究抽检地点与抽检不合格项目的关系,分析这些不合格项目的相似性关系。

(3) $p \rightarrow (L \times F \rightarrow M)$: 对于选定的抽检项目 $p \in P$, M 是抽检地点集合 L 和食品集合 F 形成的二分图的匹配。研究不同抽检地点与抽检不合格食品的关系,展示该地点范围内检出不合格项目的食品种

类、性质。

投影是沿一个维度创建数据立方体内容的聚合操作。聚合不但可以作为整体应用于集合 L 、 F 和 P , 而且可以应用于集合的子集^[17]。任务类型与基于切片的任务相对应, 包括: ① $\sum L \rightarrow (F \times P \rightarrow M)$, 忽略抽检地点因素, M 是集合 F 和集合 P 形成的二分图的匹配, 研究不同食品 and 不合格抽检项目的关系。② $\sum F \rightarrow (L \times P \rightarrow M)$, 对于所有食品一起考虑, M 是集合 L 和集合 P 形成的二分图的匹配, 研究不同地点出现抽检不合格项目的内在逻辑。③ $\sum P \rightarrow (L \times F \rightarrow M)$, 忽略所有抽检不合格项目因素, M 是集合 L 和集合 F 形成的二分图的匹配, 研究随着地点变化, 出现抽检不合格的食品种类的变化趋势。

3.2 视图设计

从任务需求出发, 构造 ExploreView 的层次图、细节描述图和关系挖掘图, 基于 Gephi、D3.js^[21] 和 Echarts.js^[22] 进行脚本编码。数据传输到请求的 JSON 输入结构中, 并且可以创建、调整、分析和下载生成的可视化效果。通过层次图, 可以对数据项有最基本的了解。通过细节描述图, 可以对食品药品监督管理局抽检数据集中的具体数值进行获取, 对用户关心并且钻取的数据关系和细节进行详细展示。通过关系挖掘图, 可以得到抽检食品与不合格抽检项目之间的关系, 探索节点间的内在关联。

3.2.1 层次图

概览视图的作用为可视化食品药品监督管理局抽检的基本情况。根据任务需求选择合适的多维模型维度, 再对每个维度进行展开。

首先展示抽检基本信息, 根据专家知识用层次树展示食品种类和抽检项目种类, 展示这两个维的层次关系, 使用户对展示对象有大概的了解。该树图可以根据需要进行展开和收缩, 当树中节点包含子节点时, 节点颜色为灰色, 否则为白色, 如图 4 所示。

3.2.2 细节描述图

细节描述图结合了多种视图, 其中包括可视化查询条件筛选设计、检测关系及频率分布图设计、时空热力图设计和检测具体信息列表等, 对用户关心并且钻取的数据关系和细节进行详细展示, 试图提供一个多尺度的编码方案来可视化用户交互时不同层次的细节。数据钻取是在分析时选择数据立方体的投影方向, 改变维的力度对数据进行层层深入的查看, 主要是在建立分析时设定钻取的维度和层次, 在查看时通过鼠标点击某个数据点时就能捕捉到下

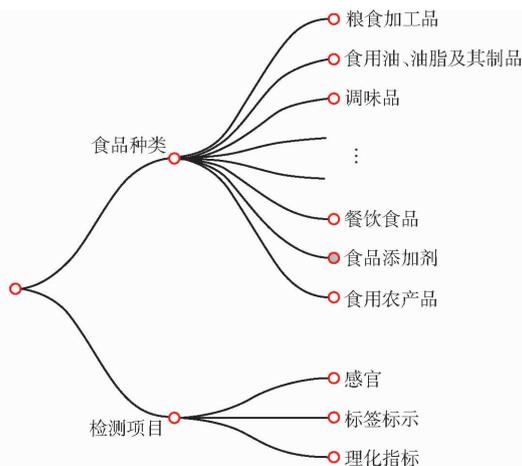


图4 层次图示意图

Fig. 4 Schematic of hierarchy

个页面。具体步骤包括:

(1) 获取用户选择的省份 l_u 并设为焦点, 将焦点作为切点产生数据立方体切片, 获取切片对应的检测报告中的两个维度, 计算相关联的二分图 G_1 , 得到集合 F_1 和 P_1 , 获取相应的食品名称和检测项目名称。

(2) 根据用户分析的目标, 连接集合 F_1 和 P_1 的相应节点, 计算得到新的二分图 G_2 , 用力导向图、圆环图等进行数据的展示。

(3) 在圆环图中, 将二分图 G_2 的食品集合 F_2 和抽检项目集合 P_2 分别可视化为不同颜色的视觉元素, 加以区分, 以更好地体现两者的关联。计算 G_2 所有的食品抽检中的不合格检测项目和检出频率, 采用强对比度的配色方案, 根据其值可视化映射为食品类型的节点颜色、抽检项目的节点颜色、连线的粗细。若某种食品检出某不合格项目, 两者之间就进行连线。所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, 如图 5a 所示。为了避免信息杂乱影响用户体验, 当数据节点超过设定值时, 抽检项目名称或食品名称将根据需要被隐藏。为了避免不同食品曲线交错复杂, 用户可以点击感兴趣的曲线, 其他线段将被隐去, 同时抽检食品与抽检项目及简单的抽检详情以浮动窗口的形式展现, 如图 5b 所示。

(4) 用户可以进行进一步分析, 通过过滤不合格抽检频率、食品名称、不合格项目类型, 显示符合要求的抽检数据, 突出显示用户关心的内容, 展示经过过滤得到的集合 F_3 和 P_3 的可视化结果, 实现交互式动态协调, 如图 5c 所示。

(5) 在 3D 柱状图中, 将二分图 G_2 的食品集合 F_2 和抽检项目集合 P_2 可视化为横纵坐标, 过滤掉不需要的集合, 分析一定范围内的食品集合 F_4 和抽检项目集合 P_4 , 并形成二分图 G_4 , 以准确地显示两者的

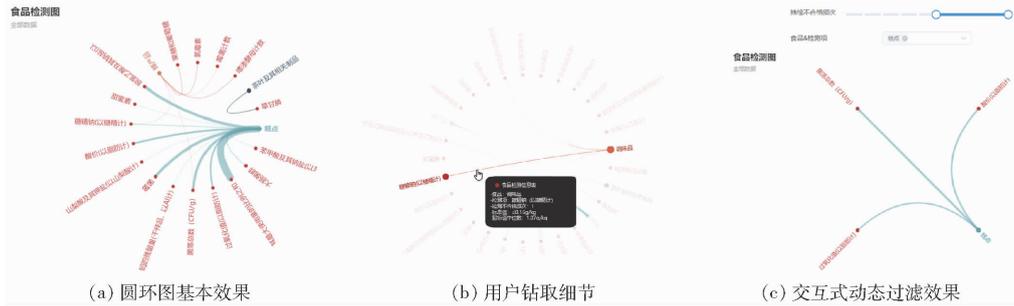


图5 圆环图

Fig. 5 Schematics of doughnut chart

数量关系,并可视化为不同高度、不同颜色的视觉元素。纵坐标 Z 表示检出频次,即所检出的不合格项目频率越高,柱状越高,颜色越向暖色调接近。用户可以进行进一步分析,通过过滤不合格抽检频率、食品名称、不合格项目类型,展示符合要求的抽检数据,展示经过过滤得到的食品集合 F_5 和抽检项目集合 P_5 的可视化结果,如图6所示。

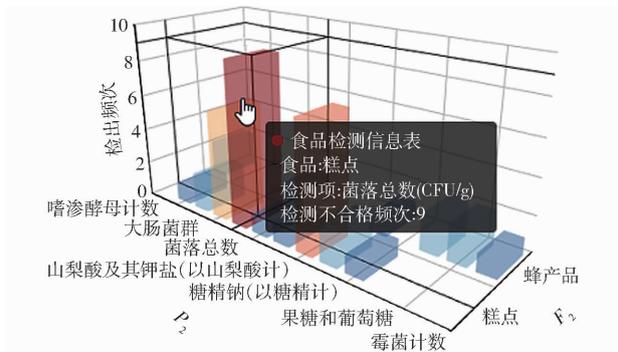


图6 3D柱状图

Fig. 6 Schematic of 3D histogram

(6) 改变用户焦点,获取用户选择的食品类别 P_n 并设为焦点,使数据立方体沿着焦点轴的方向投影,计算切片相关联的二分图 G_6 ,得到抽检地点集合 L_6 和食品集合 F_6 ,获取 L_6 集合中所有元素在集合 F_6 中对应的关联边集合,并将关联边集合中的元素数量可视化为时空热力图中的不同颜色的视觉元素。用户可以进行进一步分析,通过过滤食品名称、不合格项目类型展示符合要求的抽检数据,展示各省份特定食品类型的特定项目检出情况。若某不合格抽检项目集中出现,或按照某种规律出现,展示出异常点的特征,就可能预示着即将爆发的食品安全事件,如图7所示。

3.2.3 关系挖掘图

为了挖掘抽检食品不合格项之间的关系,探索节点间的内在关联,使用相似度计算和聚类分析数据,并基于层次扇形绑定的思想,通过热力散点同心圆图展示聚类结果。用户可通过点击节点钻取更加详细的信息。关系挖掘图的设计步骤如下:

(1) 获取抽检产品名称及其所属食品类型,可

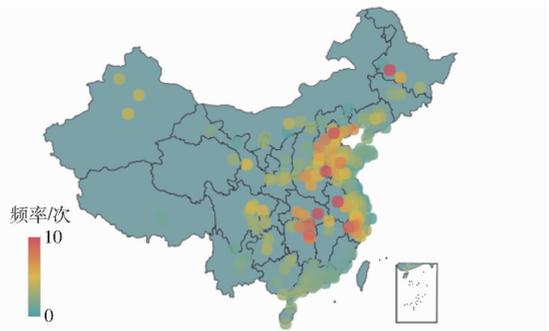


图7 时空热力图

Fig. 7 Map of space-time heat

根据食品类型进行动态过滤。

(2) 根据抽检产品的不合格项目类型和不合格项目与抽检产品的关联关系,确立食品类型数据点。首先根据《中华人民共和国食品安全法》等相关法律法规对每类食品的抽检项目进行粗分类。然后将每个抽检食品类型向量表示为

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{in}) \quad (1)$$

式中 \mathbf{x}_i ——第 i 个抽检食品类型

x_{in} ——第 i 个抽检食品与项目类别 n 的关系标识,若抽检出的不合格项目属于抽检项目类别 n ,则标识值为 1,反之为 0

抽检食品类型之间的距离计算公式表示为

$$d(\mathbf{x}_i, \mathbf{y}_j) = |x_{i1} - y_{j1}| + |x_{i2} - y_{j2}| + \dots + |x_{in} - y_{jn}| \quad (2)$$

式中 $d(\mathbf{x}_i, \mathbf{y}_j)$ ——第 i 个抽检食品与第 j 个抽检食品的距离

\mathbf{y}_j ——第 j 个抽检食品类型

y_{jn} ——第 j 个抽检食品与项目类别 n 的关系标识,若抽检出不合格项目属于抽检项目类别 n ,则标识值为 1,反之为 0

(3) 根据实际需求确定聚类个数,表示为 k 。根据抽检食品类型之间的距离进行 K-means 聚类^[23],得出聚类向量,迭代上述过程直到满足聚类终止条件。

(4) 根据上述聚类结果,所有节点组成 k 个簇。将最后的聚类结果写入可视分析数据表,作为可视

化映射的依据。

(5)用各个热力散点同心圆图展示各个簇的聚类结果。簇中的食品类型组成同心圆的外环,以扇形外环的颜色代表食品的类型,如图8a所示。

(6)抽检项目的频次和性质可视化在同心圆的内部。扇形外环代表的食品对应的检出项目类型以扇形同心圆形式按照总的检出频率由大到小由外向内排列,如图8b所示,属于同一类别的检出项目半径相同。节点的大小代表具体检出项目的检出频次,节点的颜色代表不同的检出项目,如图8c所示。

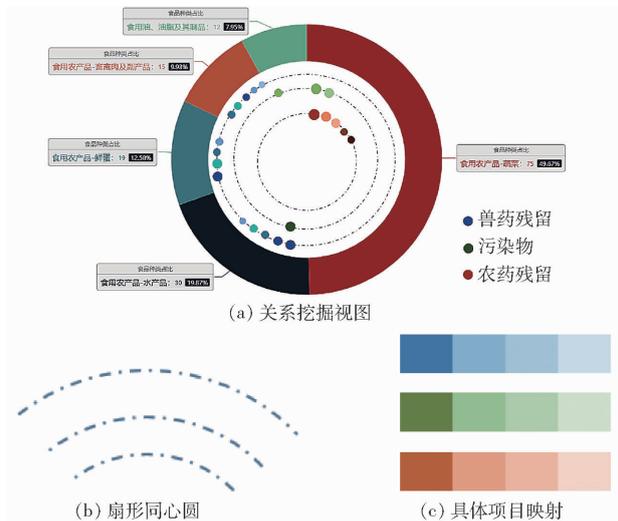


图8 关系挖掘图示意图

Fig. 8 Schematics of relationship mining map

4 分析与讨论

为验证视图的有效性,利用 ExploreView 对食品药品监督管理局发布的不合格食品抽检报告进行可视分析。

4.1 食品与抽检项目关系分析

视图提供了直观的交互功能。过滤条件选择窗口支持对一定类型、频次和时间的检测项目进行过滤。用户可以对视图进行拖动、放大或缩小,并进行多个视图的链接分析。如图9a所示,在过滤条件选择窗口,用户选择了茶叶及其相关制品、糕点、蜂产品等食品类型,选择不合格频次为3~25次。经过动态过滤,用户可通过生成的视图得到相关食品的不合格项目检出情况,并获得两者的关联关系。如图9b所示,由于节点个数小于之前设定的完全显示节点最大数(30个),所以所有的节点都进行了显示。其中,食品类型的节点(茶叶及其相关制品、糕点、调味品和蜂产品)分别为深蓝色、孔雀蓝、薄荷绿和橙色,其他抽检项目节点为红色。可以很清楚地看出,糕点类型和调味品类型的食品抽检出的不合格项目最多(连线较多),且频次高(连线相对而

言较粗)。将鼠标置于“调味品”节点上,如图9c所示,可以突出显示调味品在选定时间范围内检测出的不合格项目名称。将鼠标置于“调味品”和“糖精钠(以糖精计)”之间的连线上,可以显示两者之间的抽检关系,包括检测不合格频次、标准值和超标值中位数,如图9d所示。将鼠标置于抽检项目中的“糖精钠(以糖精计)”节点上,可以突出显示检出此种不合格项目的食品类型,即蜂产品、调味品和糕点,如图9e所示。如果仅想保留本次食品、抽检项目的计算关联边,可在过滤条件选择窗口删除多余的节点项目,并由此进行更深入的数据钻取。

在图9f中,可以看出抽检项目和食品类型之间的数量关系。可以发现,大多数抽检的检出频次在5次以下,但其中“糕点”的不合格抽检项目“防腐剂混合使用时各自用量占其最大使用量的比例之和”的柱形为红色,频次高达25次,且标准值为小于1,超标值中位数为1.5。

4.2 食品与抽检地点关系分析

在过滤条件选择窗口,选择“糕点”这一类型,选择添加剂相关的不合格抽检项目。如图9g所示,图中有标记的省份都在该食品类型中抽检出了该不合格项目,绿色节点、黄色节点至红色节点的颜色过渡代表抽检不合格频次的增加。其中红色所标记的省份河南省抽检出的频次最多,应当引起有关部门的重视。将鼠标置于节点上,可查看该省份、该食品种类、该抽检项目的具体抽检信息,比如在所选时间段内,北京市糕点食品被抽检出添加剂项目不合格的频次为2次,在全国范围内属于不合格频次较低的省份。

4.3 食品与抽检项目聚类关系分析

对所有抽检信息进行过滤、聚类,可得到不合格抽检项目与被抽检食品的相关性信息。如图9h所示,视图中展示了不合格抽检项目分类信息的层次关系,不合格抽检项目之间的关联关系以及被抽检食品与不合格抽检项目的关联关系。通过钻取点击可知,“添加剂”不合格项目包括“防腐剂混合使用时各自用量占其最大使用量的比例之和”、“脱氢乙酸及其钠盐”、“山梨酸及其钾盐”等内容。“微生物”项目的“大肠杆菌”和“菌落总数”等经常被同时检测出不合格。通过聚类可知,肉制品、调味品、饮料、方便食品、酒类在理化指标、微生物和添加剂这几类项目上经常出现不合格现象,这与它们的食品理化性质有关,也与加工过程有关,生产过程或者保存过程中的不合规操作行为具有相似性,需要根据具体情况给予更多关注。比如,调味品和肉制品频繁在添加剂项目上出现不合格现象,突出表现在



图9 案例分析效果

Fig.9 Schematics of case analysis

“防腐剂混合使用时各自用量占其最大使用量的比例之和”这一方面,反映出企业可能忽视了混合使用防腐剂时对其使用量的控制。而肉制品和饮料在微生物项目上常常不达标。因此,有关部门在抽检发现某种食品的某种项目不达标现象时,也应对聚类得到的类似食品种类提高警惕,注意潜在的共性和联系,根据实际情况重点抽检。

5 结束语

提出了一种挖掘数据关系的可视分析视图

ExploreView。在展示数据信息基本情况的同时,根据用户的实际需求进行动态交互,体现数据的属性和数据间的各种层次结构以及关联关系。用指定时间内食品药品监督管理局食品抽检不合格报告的信息数据集对本文方法进行验证,结果表明,ExploreView可以全面地展示数据基本信息,并挖掘出不合格抽检项目的细节层次关系、关联关系、潜在的相似性,以及被抽检食品与抽检项目的关联信息,实现对统计分布结果的展示和交互,为食品质量安全的监测和预警提供帮助。

参 考 文 献

[1] 陈为,沈则潜,陶焯波. 数据可视化[M]. 北京:电子工业出版社,2013.

[2] 孙扬,蒋远翔,赵翔. 网络可视化研究综述[J]. 计算机科学, 2010, 37(2):12-18.
SUN Yang,JIANG Yuanxiang,ZHAO Xiang. Survey on the research of network visualization[J]. Computer Science, 2010, 37(2): 12-18. (in Chinese)

[3] 任磊,杜一,马帅,等. 大数据可视分析综述[J]. 软件学报,2014, 25(9): 1909-1936.
REN Lei, DU Yi, MA Shuai, et al. Visual analytics towards big data[J]. Journal of Software, 2014, 25(9): 1909-1936. (in Chinese)

[4] ZHAO J, CHEVALIER F, COLLINS C, et al. Facilitating discourse analysis with interactive visualization [J]. IEEE Transactions on Visualization and Computer Graphics, 2012, 18(12): 2639-2648.

[5] COLLINS C, CARPENDALE S, PENN G. DocuBurst: visualizing document content using language structure[J]. Computer Graphics Forum, 2009, 28(3): 1039-1046.

- [6] PAULOVICH F V, MINGHIM R. Hipp: a novel hierarchical point placement strategy and its application to the exploration of document collections[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2008, 14(6): 1229 – 1236.
- [7] LAI C, ZHAO Y, YUAN X. Exploring high-dimensional data through locally enhanced projections[J]. *Journal of Visual Languages and Computing*, 2018, 48:144 – 156.
- [8] CHEN Y, ZHANG X, FENG Y, 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.
- [9] CHEN Y, DONG Y, SUN Y, et al. A multi-comparable visual analytic approach for complex hierarchical data[J]. *Journal of Visual Languages and Computing*, 2018, 47:19 – 30.
- [10] CHUI K K H, WENGER J B, COHEN S A, et al. Visual analytics for epidemiologists: understanding the interactions between age, time, and disease with multi-panel graphs[J]. *Plos One*, 2011, 6(2): e14683.
- [11] LI M, BAO Z, SELLIS T, et al. HomeSeeker: a visual analytics system of real estate data[J]. *Journal of Visual Languages and Computing*, 2018, 45: 1 – 16.
- [12] 杨信廷, 钱建平, 孙传恒, 等. 农产品及食品质量安全追溯系统关键技术研究进展[J/OL]. *农业机械学报*, 2014, 45(11): 212 – 222.
YANG Xinting, QIAN Jianping, SUN Chuanheng, et al. Key technologies for establishment agricultural products and food quality safety traceability systems[J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2014, 45(11): 212 – 222. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20141133&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2014.11.033. (in Chinese)
- [13] 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 Yi, LIU Ying, TIAN Shuai, 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. (in Chinese)
- [14] 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.
- [15] 韦津. 食品安全现状及食品安全检测方法研究[J]. *科技风*, 2018(18): 209 – 211.
- [16] 宋风平. 加强食品安全管理的对策与建议研究[J]. *食品安全导刊*, 2018(17): 75 – 76.
- [17] LI J, CHEN S, CHEN W, et al. Semantics-space-time cube: a conceptual framework for systematic analysis of texts in space and time[C]//*IEEE Transactions on Visualization and Computer Graphics*, 2018:1.
- [18] GRAY J, CHAUDHURI S, BOSWORTH A, et al. Data cube: a relational aggregation operator generalizing group-by, cross-tab, and sub-totals[J]. *Data Mining and Knowledge Discovery*, 1997, 1(1): 29 – 53.
- [19] 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, 30(1): 79 – 89.
DU Xiaomin, CHEN Yi, LI Yue. TransGraph: a transformation-based graph for analyzing relations in data set[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2018, 30(1): 79 – 89. (in Chinese)
- [20] HE X, GAO M, KAN M Y, et al. Birank: towards ranking on bipartite graphs[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(1): 57 – 71.
- [21] BOSTOCK M, OGIEVETSKY V, HEER J. D3 data-driven documents[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2301 – 2309.
- [22] LI D, MEI H, SHEN Y, et al. ECharts: a declarative framework for rapid construction of web-based visualization[J]. *Visual Informatics*, 2018, 2(2): 136 – 146.
- [23] KANUNGO T, MOUNT D M, NETANYAHU N S, et al. An efficient K-means clustering algorithm: analysis and implementation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24: 881 – 892.