

论著

2015年我国稻谷中主要元素类物质相关分析

杨大进¹,姜桥²,何来英¹,万劫³,蒋定国¹,刘卿¹,肖革新¹

(1. 国家食品安全风险评估中心,北京 100022; 2. 北京工业大学应用数理学院,北京 100124; 3. 首都经济贸易大学,北京 100070)

摘要:目的 通过对稻谷中主要元素进行相关关系分析,为今后开展食品安全风险监测数据的挖掘利用提供借鉴。方法 采用贝叶斯网络挖掘稻谷中主要元素的相关关系,并根据结果对元素进行分组,并对分组后的元素进行典型相关分析,分析过程采用R统计软件实现。结果 通过贝叶斯网络分析,确定稻谷中有两组元素含量之间有直接或间接的相互影响关系。根据贝叶斯网络分析结果,对两组元素进行典型相关分析,第一典型相关系数为0.84,第一典型变量中钒含量的系数最高,且呈现负相关关系,铝和镉含量的系数较高,且与铝的含量呈负相关,而与镉的含量呈正相关。结论 应用贝叶斯网络和典型相关分析方法可以更好地揭示稻谷中主要元素含量之间的相关关系。

关键词:稻谷;元素;相关关系;贝叶斯网络;典型相关分析;食品污染物

中图分类号:R155 文献标识码:A 文章编号:1004-8456(2017)02-0136-04

DOI:10.13590/j.cjfh.2017.02.004

Correlation analysis of major elements of paddy in China in 2015

YANG Da-jin¹, JIANG Qiao², HE Lai-ying¹, WAN Jie³, JIANG Ding-guo¹, LIU Qing¹, XIAO Ge-xin¹

(1. China National Center for Food Safety Risk Assessment, Beijing 100022, China;

2. College of Applied Sciences, Beijing University of Technology, Beijing 100124, China;

3. Capital University of Economics and Business, Beijing 100070, China)

Abstract: Objective Through the correlation analysis of the main elements of paddy in China in 2015, this paper could provide reference for the data mining of food safety risk monitoring result. **Methods** The correlation of major elements in paddy were analyzed by Bayesian network, and the elements were grouped according to the results. Canonical correlation analysis between the grouped elements was carried out by R statistics software. **Results** Bayesian network found that there were direct or indirect correlations among the contents of two groups. The first typical correlation coefficient between the two groups of elements was 0.84 according to the canonical correlation analysis. The negative coefficient of V in the first typical variable was the highest, and the coefficients of Al and TI are high, and negatively correlated with the Al content, while positively correlated with the TI content. **Conclusion** Bayesian network and canonical correlation analysis could analyze the correlation among the major elements in paddy.

Key words: Paddy; element; correlation analyze; Bayesian network; canonical correlation analysis; food contaminant

随着现代工农业的发展,稻谷从被污染的空气、水和土壤中吸收并富集了大量的有毒有害物质,尤其是重金属。大量重金属进入人体后会破坏人体酶的正常活动,从而对人体生理造成严重影响,甚至导致死亡^[1-5]。在食品污染数据分析工作中,经常会发现在某一类食品中某一种污染物含量升高时,另一种污染物含量也会随之升高或降低,这种现象说明这两种污染物的含量可能存在某种

相关关系^[6-7]。研究稻谷中不同元素含量的相互关系,有助于发现不同元素在稻谷中的关联关系,通过进一步探索这些元素间相关关系的具体表现形式及形成原因,为土壤治理、种植方案、饮食管理和政府政策的制定提供依据。目前,我国食品安全风险监测部门已收集了大量食品污染物数据,这些数据的充分利用将能够为我国的食品安全工作提供有力支持。

收稿日期:2016-12-27

作者简介:杨大进 男 研究员 研究方向为食品安全监测及预警

E-mail: yangdajin@cfsa.net.cn

通信作者:肖革新 男 副研究员 研究方向为空间流行病学

E-mail: xiaogexin@cfsa.net.cn

1 材料与方法

1.1 材料

2015年食品安全风险监测工作中,对全国18个地市共采集到的2363份稻谷样品中14种主要

元素含量进行了检测,包括:钡、钒、镉、锂、铝、锰、铅、铈、锑、铜、硒、总铬、总汞、总砷。

1.2 方法

1.2.1 贝叶斯网络

贝叶斯网络算法可以得出稻谷中污染物含量之间相互影响的路径,从而发现多项污染物含量之间复杂的相关关系^[8-10]。

贝叶斯网络构建一个图模型,图中的节点与变量一一对应,并且满足:

$$p(X_1, X_2, \dots, X_n) = \prod_{i=1}^n p(X_i | Pa_i)$$

其中, X_1, X_2, \dots, X_n 为随机变量, $p(0)$ 表示概率密度函数, Pa_i 表示 X_i 的父节点集合。构造上述图模型的问题可以转化为寻找每个随机变量的马尔科夫毯的问题。随机变量 X 的马尔科夫毯 $B(X)$ 定义为: $B(X)$ 为全体随机变量集合 V 的一个子集,满足:对任意 $Y \in V - B(X) - \{X\}$, 有:当 $B(X)$ 给定的条件下, X 与 Y 独立。

寻找 $B(X)$ 可以通过 grow-shrink 算法实现。grow-shrink 算法步骤如下:

- ①初始化集合 S , 置 S 为空集;
- ②growing 步: 只要当 $\exists Y \in V - \{X\}$ 满足当 S 给定的条件下 Y 与 X 不独立, 则置 S 为 $S \cup Y$;
- ③shrinking 步: 只要当 $\exists Y \in S$ 满足当 $S - \{Y\}$ 给定的条件下 Y 与 X 独立, 则置 S 为 $S - \{Y\}$;
- ④输出: 置 $B(X)$ 为 S 。

通过变量的马尔科夫毯, 可以构造出贝叶斯网络, 在贝叶斯网络中, 若节点 A 到 B 有连边则 A 与 B 有直接的相关性, 若 A 到 B 没有直接的连边但有通路, 则 A 与 B 之间有间接的相关性^[10-12]。

1.2.2 典型相关分析

典型相关分析是利用综合变量来反映两组指标之间的整体相关性的多元统计分析方法。它的基本原理是为了从总体上把握两组指标之间的相关关系, 分别在两组变量中提取有代表性的两个综合变量 U 和 V (分别为两组变量中各变量的线性组合), 利用这两个综合变量之间的相关关系来反映两组指标之间的整体相关性^[13-15]。

一般地, 假设有一组变量 X_1, X_2, \dots, X_p 与另一组变量 Y_1, Y_2, \dots, Y_q , 当 $p = q = 1$ 时, 就是研究两个变量 X 与 Y 之间的简单相关关系, 定义为:

$$\rho_{xy} = \frac{Cov(X, Y)}{\sqrt{Var(X)} \sqrt{Var(Y)}}$$

当 p 和 q 均小于 1 时, 利用主成分分析的思想, 可以把多个变量与多个变量之间的相关化化为两

个新的综合变量之间的相关。也就是做两组变量的线性组合, 即

$$U = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_p X_p = \alpha X$$

$$V = \beta_1 Y_1 + \beta_2 Y_2 + \dots + \beta_q Y_q = \beta Y$$

其中, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_p)$ 和 $\beta = (\beta_1, \beta_2, \dots, \beta_q)$ 为任意非零向量, 将研究两组变量之间的问题化为研究两个变量 U 和 V 之间的相关问题, 希望寻求 α, β 使 U, V 之间最大可能的相关。利用 R 软件中的“cancor()”函数可以实现典型相关分析。

2 结果

2.1 贝叶斯网络分析

采用我国 18 个地市稻谷样品中 14 种元素的含量数据, 利用 R 软件编程进行贝叶斯网络训练, 算法采用 grow-shrink, 通过普通相关系数检验条件独立, 得到相应关系网络, 见图 1。

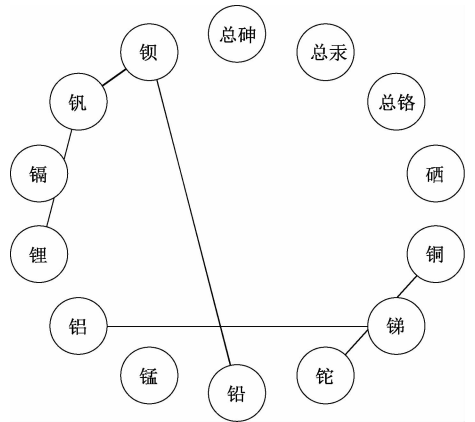


图 1 通过稻谷中金属含量数据得出的贝叶斯网络图
Figure 1 Bayes network got from the amount data of metal elements of paddies

贝叶斯网络揭示出各元素间相互影响的传递关系和路径。图中有直接连边的元素表示有直接的相互影响, 间接相连的元素虽然没有直接影响, 但会通过其他中间元素间接相互影响。

通过贝叶斯网络分析发现在我国 18 个地市的稻谷样品中锂、镉、钒、钡、铅含量之间存在直接或间接的相互影响关系。其中, 稻谷中钡含量和钒含量、钡含量和铅含量之间均有直接的相互影响关系, 根据元素化学及相关矿物学研究^[16], 钒与钡、铅与钡之间通常会呈现伴生关系, 在含量上存在高度的相关性。由此可知贝叶斯网络结果与公认的研究结果存在一致性, 贝叶斯网络结果具有一定的参考价值; 但相关研究提示有些元素在自然环境中没有明显的相关性, 贝叶斯网络却显示其在稻谷中的含量有高度相关性时, 则说明这些对应元素间的相关性可能是人为造成的, 如镉是

提取锌的副产品,电镀工业、塑胶剂制造、合金制造及焊料、染料和涂料加工会同时产生镉和锌,造成镉和锌含量的高度相关性。当贝叶斯网络发现可能人为造成元素含量之间的高度相关性时,则需进一步调查研究,以早日发现该元素的污染源头,并进行相关治理。

同理由图1可知,稻谷中另一组主要元素铝、镉、铜、铊含量之间存在直接或间接的相关关系,但两组元素中任何单个元素之间没有直接或间接关系。

2.2 典型相关分析

根据贝叶斯网络对我国稻谷中主要元素的分组情况,利用R语言中的“cancor()”函数对稻谷中主要元素组进行典型相关分析,其部分结果见表1。

表1 典型相关变量的相关系数

Table 1 Correlation coefficients of typical correlation variables

序号	相关系数
1	0.844 97
2	0.585 29
3	0.412 62
4	0.236 51

相关分析的目的是利用降维的方法来处理数据,因此首先需要作典型相关系数的显著性检验来判断选择多少对典型变量^[17]。若认为典型相关系数 $\rho_k = 0$,则就不必考虑第 k 对典型变量。通过R语言程序计算,结果表明只需一对典型变量。

用 $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9$ 分别代表数据标准化后的稻谷中钒、钨、镉、锂、铅、铝、铊、镉、铜的含量。由于只有第一典型相关变量是显著的,所以只需考虑第一组典型相关变量:

$$U_1 = 0.037X_1 - 0.230X_2 - 0.042X_3 - 0.034X_4 + 0.023X_5$$

$$V_1 = -0.418X_6 + 0.119X_7 + 0.402X_8 + 0.109X_9$$

从第一组典型相关方程,可以看出第一典型变量 U_1 中钒含量(X_2)的系数最高,且与 U_1 呈负相关关系,因此钒含量在第一典型变量 U_1 中发挥主导作用。第一典型变量 V_1 中铝含量(X_6)的系数最高,且与 V_1 的含量呈负相关,因此铝含量在第一典型变量 V_1 中发挥主导作用;同时镉含量(X_8)在第一典型变量 V_1 中发挥较重要作用,典型载荷为0.402。其余元素含量对典型变量的贡献不明显。

同时第一典型相关系数达到0.84,表明我国稻谷中第一组主要元素含量和第二组主要元素含量有较高的相关关系,即稻谷中钒含量与镉含量呈负相关关系,表现为拮抗现象;稻谷中钒含量与铝含量呈正相关关系,表现为协同现象。根据该结论,研究人员可进一步研究稻谷对钒、镉和铝的吸收、

富集机理以及钒、镉、铝在植物体内的两两相互作用,进而探究是否能利用稻谷中金属镉的作用来降低金属铝和镉的含量,或根据稻谷中钒含量与铝含量的正相关关系,寻找同时造成钒和铝增高的源头,并进行针对性治理。

3 讨论

在分析食品中污染物含量的相关关系时,本研究主要采用了贝叶斯网络分析和典型相关分析,其中贝叶斯网络可以发现不同污染物含量之间直接或间接相互影响的关系,典型相关分析能分析不同组别的污染物之间的整体相关关系,从而多角度剖析稻谷中污染物含量的直接或间接关系。

贝叶斯网络运用图形的方式揭示了变量或数据之间复杂的依赖或相互影响关系,它是对数据对象关系的高度抽象与概括,将变量之间潜在的关联性用清晰明了、自然直观的图形建模表示出来^[18]。贝叶斯网络运用了严密的数学推理方法,结果和结论可信度强,便于解释和理解;但是贝叶斯网络只能解释变量之间存在相互影响关系,却无法显示变量间具体的相关关系,如是否呈正相关或负相关,以及相互影响的强度大小。对此可进行简单相关分析以做进一步具体研究,如皮尔森相关系数等。此外贝叶斯网络的算法相对较为复杂,需要较大的样本量,但是贝叶斯网络能够很好地解释普通相关分析中伪相关现象的出现。

典型相关分析是多元统计分析的一个重要研究课题,它借助主成分的思想,用少数几对综合变量来反映两组变量间的线性相关关系。典型相关关系目前已经在众多领域的相关分析和预测分析中得到广泛应用,但鲜有用于研究稻谷中多种元素含量的相关分析。本研究在贝叶斯网络分析基础上进行典型相关分析,进一步确定了稻谷中两组元素之间具体的相关类型和相关强度。两种方法从不同角度全面分析了稻谷中多种元素含量之间的相关关系,为污染溯源提供了针对性的线索,为相关结果的分析提供了新思路。

参考文献

- [1] 王豫. 浅谈食品中重金属对人体的危害及预防[J]. 青海农技推广, 2010(4): 8-9, 16.
- [2] 于艳秋, 张笑竹, 刘娟. 重金属污染对人体的危害[J]. 企业标准化, 2008(1): 16.
- [3] 游勇, 鞠荣. 重金属对食品的污染及其危害[J]. 环境, 2007(2): 102-103.
- [4] 李爱琴, 王阳峰, 杨珊娇. 浅谈重金属污染对健康的危害[J]. 河南机电高等专科学校学报, 2005, 13(4): 49-50.
- [5] 李其林, 刘光德, 黄韵, 等. 蔬菜中重金属特征研究[J]. 中国

- 农学通报,2004,20(3):40-44.
- [6] JINADASA K B P N, MILHAM P J, HAWKINS C A, et al. Survey of cadmium levels in vegetables and soils of greater Sydney, Australia[J]. Journal of Environmental Quality, 1997, 26(4): 924-933.
- [7] 魏世强,李伟,王定勇.重庆市土壤与粮食作物重金属污染评价及相关规律研究[C].土壤资源持续利用与生态环境安全学术研讨会,广州,2009.
- [8] 慕春棣,戴剑彬,叶俊.用于数据挖掘的贝叶斯网络[J].软件学报,2000,11(5):660-666.
- [9] 冀俊忠,刘椿年,沙志强.贝叶斯网模型的学习、推理和应用[J].计算机工程与应用,2003,39(5):24-27,47.
- [10] 胡玉胜,涂序彦,崔晓瑜,等.基于贝叶斯网络的不确定性知识的推理方法[J].计算机集成制造系统,2001,7(12):65-68.
- [11] 王双成,苑森森.具有丢失数据的贝叶斯网络结构学习研究[J].软件学报,2004,15(7):1042-1048.
- [12] 林士敏,田凤占,陆玉昌.贝叶斯网络的建造及其在数据挖掘中的应用[J].清华大学学报(自然科学版),2001,41(1):49-52.
- [13] 沈家芬,张凌,莫测辉,等.广州市空气污染物和气象要素的主成分与典型相关分析[J].生态环境,2006,15(5):1018-1023.
- [14] 孙权森,曾生根,王平安,等.典型相关分析的理论及其在特征融合中的应用[J].计算机学报,2005,28(9):1524-1533.
- [15] MALEC L. Some remarks on the functional relation between canonical correlation analysis and partial least squares [J]. Journal of Statistical Computation and Simulation, 2015, 86:2379-2391.
- [16] 陕西省地质局西安实验室钒钼铜矿研究组,中国地质科学院地矿所钒钼铜矿研究组.钒钼铜矿的矿物学及其晶体结构[J].地质学报,1977(1):67-83.
- [17] 傅德印,黄健.典型相关分析中的统计检验问题[J].统计研究,2008,25(7):110-112.
- [18] 黄解军.贝叶斯网络结构学习及其在数据挖掘中的应用研究[D].武汉:武汉大学,2005.

· 资讯 ·

解读《关于食品添加剂新品种氨基乙酸(羟基乙腈法)等的公告(2017年第3号)》

一、氨基乙酸(羟基乙腈法)

(一)背景资料。氨基乙酸(羟基乙腈法)的分子式是 $C_2H_5NO_2$ 。欧盟委员会、美国食品药品监督管理局、日本厚生劳动省等允许其作为食品添加剂用于食品。

(二)工艺必要性。氨基乙酸作为增味剂和食品用香料已列入《食品安全国家标准 食品添加剂使用标准》(GB 2760),本次申请的氨基乙酸(羟基乙腈法)为新的生产工艺生产,其使用量和使用范围与已批准的氨基乙酸相同,用于预制肉制品(食品类别 08.02)、熟肉制品(食品类别 08.03)、调味品(食品类别 12.0)、果蔬汁(浆)类饮料(食品类别 14.02.03)、植物蛋白饮料(食品类别 14.03.02)、作为食品用香料用于各类食品(GB 2760 表 B.1 食品类别除外),改善食品的味道。其质量规格按照公告的相关内容执行。

二、乙基芳樟基醚

(一)背景资料。乙基芳樟基醚的分子式是 $C_{12}H_{22}O$ 。联合国粮农组织/世界卫生组织食品添加剂联合专家委员会、欧盟委员会、美国食用香料和提取物制造者协会等允许其作为食品用香料在各类食品中按生产需要适量使用。

(二)工艺必要性。该物质配制成食品用香精后用于各类食品(GB 2760 表 B.1 食品类别除外),改善食品的味道。其质量规格按照公告的相关内容执行。

三、β-胡萝卜素

(一)背景资料。β-胡萝卜素作为食品添加剂,已列入《食品安全国家标准 食品添加剂使用标准》(GB 2760),允许用于调制乳、风味发酵乳、干酪类似品、脂肪类甜品、水果罐头、果酱、蜜饯凉果、干制蔬菜、腌渍的蔬菜、糖果、方便米面制品、熟肉制品等食品类别,本次申请其使用范围扩大到调理肉制品(生肉添加调理料)(食品类别 08.02.01)。国际食品法典委员会、欧盟委员会、美国食品药品监督管理局、澳大利亚和新西兰食品标准局等允许其作为着色剂用于食品。

(二)工艺必要性。该物质作为着色剂用于调理肉制品(生肉添加调理料)(食品类别 08.02.01),调节产品色泽。其质量规格应执行《食品添加剂 β-胡萝卜素》(GB 8821—2011)或《食品添加剂 β-胡萝卜素(发酵法)》(GB 28310—2012)或国家卫生计生委 2012 年第 6 号公告的相关内容执行。

(相关链接:<http://www.nhfpc.gov.cn/sps/s3586/201703/a41bad69ee9248fcbddd9aea7d03cabe.shtml>)