

论著

压缩感知方法在食品安全风险监测中的应用

姜桥^{1,2}, 李宁¹, 何来英¹, 贾金柱³, 耿直³, 万劼⁴, 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. 北京大学, 北京 100871; 4. 首都经济贸易大学, 北京 100070)

摘要:目的 压缩感知方法通过混样过程来减少样品的检测次数,从而提高检测效率,降低检测成本,缩短检测时间。方法 利用食品中污染物超标数据稀疏性的特点,将压缩感知方法应用于食品安全风险监测的样品检测。该理论的核心思想是通过混合待检测的样品,得到远少于原样品数的检测次数,然后根据相应重构算法由测量值重构原始数据。该算法可采用 R 统计软件实现。结果 用压缩感知方法重构 125 份原始样品的检测值,误差平方和为 $3.782\ 652 \times 10^{-29}$,其中原始样品中 117 份低于检出限的样品全部精准重构,高于检出限的 8 份样品压缩感知重构值稍稍大于真实值,但误差极小,可以忽略不计。结论 压缩感知方法可以通过混合样品来减少样品的检测次数,并可由少数检测值重构每一个原始样品的食品污染物含量。

关键词:压缩感知;混样;食品安全;检测;稀疏性;R 统计;重构

中图分类号:R155 文献标识码:A 文章编号:1004-8456(2016)06-0692-03

DOI:10.13590/j.cjfh.2016.06.002

Application of compressed sensing method in food safety risk monitoring

JIANG Qiao, LI Ning, HE Lai-ying, JIA Jin-zhu, GENG Zhi, WAN Jie, XIAO Ge-xin

(China National Center for Food Safety Risk Assessment, Beijing 100022, China)

Abstract: Objective The compression sensing method reduces the times of sample inspection by sample mixing process, thus improving the sampling efficiency, reducing the sampling cost and shortening the sampling time. **Methods** Based on the characteristics of sparse data, the compression sensing method was applied to the food safety risk monitoring sample detection. The mixed samples were detected, and the original data was then reconstructed according to the corresponding algorithm with R statistical software. **Results** The detection values of 125 original samples were reconstructed using the compression sensing method, and the sum of squared errors was $3.782\ 652 \times 10^{-29}$. There were 117 original samples lower than the detection limit were accurately reconstructed. The 8 samples which were higher than the detection limit were slightly larger than the real value, but the error is very small and can be neglected. **Conclusion** Compression sensing methods could reduce the testing number of samples by mixing samples and reconstruct the contaminant content of each original sample from several mixed samples.

Key words: Compressed sensing; mixing samples; food safety; test; sparseness; R statistics; refactoring

食品安全风险监测是保障食品安全的有效措施,工作量巨大。某些污染物的检测过程复杂,设备昂贵,在有限资金和时间内检测样品过少而无法得到真正的超标率,可能导致有害食品流入市场。如何设计高效的食品安全检测方案,既减少工作量,

又保证得到的结果真实可靠成为一大难题^[1]。实际上某些污染物在食品中含量较低,通常大多数检测值低于检出限(LOD)或国家限量^[2],从而采集到的数据具有极高的稀疏性。在 2004 年,由 Donoho^[3]、Candes 等^[4-5]提出的压缩感知(compressed sensing,简称 CS)理论,为数据采集技术带来了革命性的突破,其核心思想是将压缩和采样合并进行,首先采集数据的非自适应线性投影(测量值),然后根据相应重构算法由测量值重构原始数据^[6-8]。压缩感知的优点在于检测样品次数远远小于传统方法的检测次数,以较少检测次数,通过计算就可以得到样品的超标率、检出率等指标,从而大大提高了检测效率。本文通过分析实际数据,讨论压缩感知应用于食品安全风险监测的可行

收稿日期:2016-11-23

基金项目:国家科技支撑计划课题:基于电子溯源的食品安全风险评估关键技术研究与应用(2015BAK3604);国家卫生计生委食品司委托课题:食品安全风险监测结果分析报告

作者简介:姜桥 女 研究生 研究方向为统计及食品安全风险分析
E-mail:qiaojiang_gogogo@163.com

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

性,并对检测效果进行评估。

1 材料与amp;方法

1.1 数据来源

2015年湖南省葡萄酒中安赛蜜的检测数据,单位为g/kg,共计125条。

1.2 方法

压缩感知理论指出:如果信号在某个变换域是稀疏的或可压缩的,就可以利用一个与变换基不相关的观测矩阵将变换得到的高维信号投影到一个低维空间上,根据这些少量的观测值,通过求解凸优化问题就可以实现信号的精确重构^[9]。压缩感知理论主要包括:信号的稀疏表示、编码测量和重构算法三个方面^[10]。

CS理论的前提是稀疏性和不相关性^[11]。信号 x (设长度为 N)的稀疏表示就是将信号投影到正交变换基时,绝大部分交换系数的绝对值很小,只有 K ($K \ll N$)个是非零值,即为 K -稀疏信号。

在编码测量中,用一个与变换矩阵不相关的 $M \times N$ [$K \lg(N/K) < M \ll N$]测量矩阵 ϕ 对信号进行线性投影,得到线性测量值 y :

$$\phi x = y \tag{1}$$

测量值 y 是一个 $M \times 1$ 矩阵,从而使测量对象从 N 维降到 M 维。

最后,运用重构算法由测量值 y 及投影矩阵 ϕ 重构原始信号 x 。这类求逆问题可以通过求解最小0-范数从 y 中恢复 x :

$$\hat{S} = \underset{S'}{\operatorname{argmin}} \| S' \|_0 \tag{2}$$

$$\text{s. t } \phi \hat{S} = y$$

可以证明只要利用 $M = K + 1$ 个独立同分布的高斯测量值就可以用最小0-范数法以高概率重构 K -稀疏信号,但(2)式的求解复杂度、稳定性差而且是一个很难找到多项式时间算法(NP-hard)的问题^[12]。可以证明用 L_1 范数代替(2)式中 L_0 范数会得到同样的解^[13]:

$$\hat{S} = \underset{S'}{\operatorname{argmin}} \| S' \|_1 \tag{3}$$

$$\text{s. t } \phi \hat{S} = y$$

只要利用 $M \geq K \lg(N/K)$ 个独立同分布的高斯测量值就可以以高概率重构 K -稀疏信号。

在食品检测中,假设原始检测样品数为 p ,通过计算机产生 $m \times p$ 的0-1随机矩阵, m 为实际测量次数,其中 m 远远小于 p 。将随机矩阵的每一行与原始样品对应,将随机矩阵中1对应的样品取出,充分混合后进行检测,得到一个检测值 y_i ,如此进行 m 次混和样品,得到 m 个检测值。然后通过求解凸优化问题由0-1随机矩阵和 m 个检测值重构原始 p 个

样品的检测值。具体混样、检测方案可表示如(4),0-1随机矩阵的每一行与原样品的列向量相乘,1代表取样,0为不取。0-1随机矩阵与原样品矩阵相乘得到 m 个检测值。

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 1 & 0 & 0 & 1 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_p \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \tag{4}$$

2 结果与分析

本文采用2015年湖南省葡萄酒中安赛蜜的检测数据进行计算。

共125份样品数据,安赛蜜的检出限为0.05g/kg,其中有8份样品安赛蜜的检测值高于检出限,117份样品未检出安赛蜜。为拉大检出值与未检出值的差距,将未检出值定为0^[14],则125份样品中安赛蜜含量的检测值分布如图1。

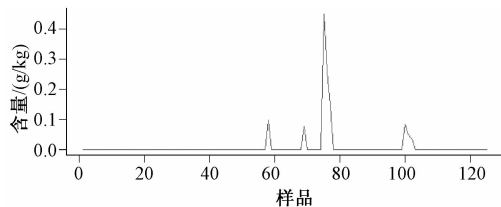


图1 葡萄酒中安赛蜜检测值

Figure 1 Detection values of acesulfame potassium in wine

图1中横坐标代表125份样品,纵坐标代表每份样品中安赛蜜的含量,可以看出大部分样品数据(安赛蜜含量)比较小,只有少数几份样品数据比较大,因此该数据符合“稀疏性”的特征,且在抽样过程中,每个样品都是独立抽取的,符合“不相关性”的要求,因此可以采用压缩感知方法进行混样检测。

在该例中强信号为8,应选混合次数 M 需满足 $K \lg(N/K) < M \ll N$,即 $22 < M \ll 125$,故选取30次。用R软件随机生成一个 30×125 的0-1矩阵,对样品进行混合,得到30份混合样品的安赛蜜检测值。采用求解 L_1 范数的优化方法反推原始125份样品中安赛蜜的含量。125份样品的实测值与重构值的分布如图2。图2中实线代表样品中安赛蜜的实测值,点代表压缩感知混合方法重构得出的安赛蜜值,可以看出压缩感知方法能很好的重构原始样品中安赛蜜的含量。

安赛蜜检测值低于检出限的样品全部精准重构,安赛蜜高于检出限的样品实测值与该方法重构值对比见表1。

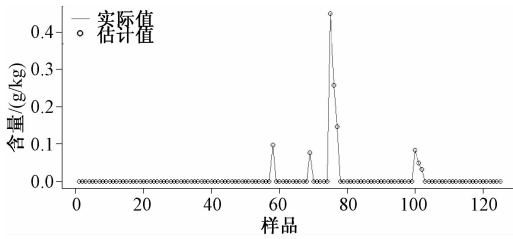


图2 压缩感知计算值与实测值对比图

Figure 2 Comparison chart of the compressed sensing values and the real values

表1 样品中安赛蜜实测值与压缩感知重构值对比

Table 1 Comparisons of true values and compressed sensing reconstructed values of sample contaminants

实测值	重构值	误差(重构值-实测值)
0.096 988 14	0.096 988 14	$4.371\ 503 \times 10^{-15}$
0.077 008 36	0.077 008 36	$6.106\ 227 \times 10^{-16}$
0.449 660 89	0.449 660 89	$8.326\ 673 \times 10^{-16}$
0.258 262 60	0.258 262 60	$1.665\ 335 \times 10^{-15}$
0.146 940 51	0.146 940 51	$-1.665\ 335 \times 10^{-16}$
0.082 934 56	0.082 934 56	$2.359\ 224 \times 10^{-16}$
0.049 130 89	0.049 130 89	$3.573\ 530 \times 10^{-15}$
0.032 451 07	0.032 451 07	$1.422\ 473 \times 10^{-15}$

从表1可以看出,高于检出限的样品污染物实测值与压缩感知方法重构值误差极小,约为检出限浓度的 $1/10^{13}$,若重构误差以误差平方和表示,为 $3.782\ 652 \times 10^{-29}$,可以忽略不计,因此压缩感知方法能精确重构原始样品中安赛蜜的含量。

3 小结

压缩感知理论的诞生在学术界引起了极大的反响,已经对理论数学、计算数学、计算科学、概率论、信息论、信号处理、电子工程、光学工程等诸多领域产生了重要影响。采用压缩感知方法的前提是稀疏性和不相关性,它要求待测样品中被检测物质的历史检测值大多数都比较小,只有少数几个样品中被检测物质的含量较大,并且该规律会长久维持下去。另外该样品中被检测物质的检测方法主要是通过液体的检测实现的,如液体类食品或通过液体浸泡提取被检测物质的食品(如检测蔬菜中的污染物)。可以根据产生的随机矩阵混合样品并进行检测,得到至少 $K \lg(N/K)$ 个检测值,并通过求解

凸优化问题估算原样品的污染物含量,或直接计算检出率、合格率等指标。在该理论框架下,压缩感知方法通过混样来减少检测次数,从而提高了检测效率,节省了检测费用和时间。压缩感知方法在需要检测大量样品的工作中倍显优势,如海关食品安全检测、工厂食品质量控制等。但压缩感知方法在食品安全检测中的应用目前还处于理论阶段,需各单位结合具体实例进行多次重构验证,寻找达到最优重构结果的最少次数,将理论融入实际工作中。

参考文献

- [1] 王凤娇,祝旭.食品安全检测存在的问题及改善措施[J].现代食品,2016,3(5):11-12.
- [2] 王绪卿,吴永宁,陈君石.食品污染监测低水平数据处理问题[J].中华预防医学杂志,2002,36(4):278-279.
- [3] Donoho D L. Compressed sensing[J]. IEEE Transactions on Information Theory,2006,52(4):1289-1306.
- [4] Candes E J,Wakin M B. An introduction to compressive sampling[J]. IEEE Signal Processing Magazine,2008,25(2):21-30.
- [5] Candes E J,Romberg J,Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information[J]. IEEE Transactions on Information Theory,2006,52(2):489-509.
- [6] PENG Y,Ganesh A,Wright J,et al. RASL: robust alignment by sparse and low-rank decomposition for linearly correlated images[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2012,34(11):2233-2246.
- [7] Tsaig Y,Donoho D L. Extensions of compressed sensing[J]. Signal Processing,2006,86(3):549-571.
- [8] Romberg J. Imaging via compressive sampling[J]. IEEE Signal Processing Magazine,2008,25(2):14-20.
- [9] 王东梅,侯晓赞.压缩感知理论及其应用前景[J].中国新通信,2010(21):71-73.
- [10] 石光明,刘丹华,高大化,等.压缩感知理论及其研究进展[J].电子学报,2009,37(5):1070-1081.
- [11] 沈爱民.压缩感知:通信与信号处理领域中的机遇与挑战[M].北京:中国科学技术出版社,2013.
- [12] 文方青,张弓,陶宇,等.面向低信噪比的自适应压缩感知方法[J].物理学报,2015,64(8):084301.
- [13] 李卓凡,闫敬文.压缩感知及应用[J].微计算机应用,2010,31(3):12-16.
- [14] WHO. Guidelines for the study of dietary intakes of chemical contaminants[J]. WHO Offset Publication,1985(87):1-102.